1996

Application Of Expert Systems In Analytical And Environmental Chemistry

Qiwei Zhu

APPLICATION OF EXPERT SYSTEMS IN ANALYTICAL AND ENVIRONMENTAL CHEMISTRY

by

Qiwei Zhu

Department of Chemistry

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario

September 1995

ISBN 0-612-09892-3

ABSTRACT

Expert systems are knowledge based computer programs that offer the possibility of combining theories with heuristic expertise, and are therefore capable of solving domain-specific problems. However, little is known about how human knowledge can be effectively and precisely transferred into computer programs. In addition, the lack of a seamless man-machine interface also causes reduced general acceptance of expert systems. In this thesis, the process of knowledge acquisition and the role of the user interface in expert systems are investigated. Results are presented that describe the development of two prototypic expert systems: SPILLexpert and GSMSdiagnosis.

The knowledge domain matrix (KDM) is an alternative knowledge encoding mechanism that provides a causal knowledge model to be followed in the phases of knowledge extraction and interpretation. The results from the causal analysis are arranged into knowledge tables from which representation to a computer data structure, in this work production rules, can be easily achieved. The KDM approach is flexible and effective in the subsequent development phases when the knowledge base is under frequent modification and/or expansion. Knowledge bases developed in this way are portable and can be readily transferred to other expert system shells.

The ACexpert graphical user interface was developed in the Microsoft Windows 3.1 environment. It provides an intuitive access to these systems through the use of graphics and pull down menus, which can be activated by either point-and-click or key stroke. This user interface has been implemented in both the SPILLexpert and GCMSdiagnosis programs.

SPILLexpert is the first expert system application of its kind developed to assist in the response to chemical spills. The importance of this work is two fold. First, an expert system holding heuristics is interfaced with a database structure that retains factual information. Second, the finished program provides a

systematic and integrated approach to the development of the most appropriate response to be chosen following an environmental emergency.

GCMSdiagnosis is a diagnostic expert system for trouble-shooting operational problems with the quadrupole ion-trap GC-MS instrument. This work demonstrates the possibility of developing portable expert systems to enhance the 'intelligence' of modern analytical instruments.

DEDICATION

I dedicate this thesis to my dear wife and son. For all their love, understanding, and support, this thesis is as much theirs as it is mine.

Thank you.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to all the people who had supported me in my five years of graduate studies in Western, without their help this work would not have been possible. First of all, I would like to thank Dr. Martin Stillman, my supervisor, for his patience, guidance, and encouragement. His view of the potential of expert systems in chemistry has constantly inspired me to pursue my own career in the field. Special thanks to Dr. Ray March and members in his research group, I had a wonderful opportunity to work with them and successfully completed the joint project. I would also like to thank my lab-mates for their friendship, help and discussions: Dr. Devon Cancilla, John Dixon, Hai Du, Dr. Anna Rae Green, Dr. Ziqi Gui, Dr. Guosheng Huang, Dr. Sharbari Lahiri, Dr. Wuhua Lu, Xin Li, Dr. Shengping Ma, Dr John Mack, Dr. Ed Ough, and Dr. Anthony Presta. I am especially grateful to Dr. Lahiri, Mr. Du, Drs. Huang, Ma and Cancilla, with whom I had worked closely and had many valuable and intriguing discussions. I also appreciate the help and kindness offered by all the staff in the Chemistry Department, in particular, the staff in secretarial, electronic-shop, chemistry store, and third-year analytical laboratory.

I wish to acknowledge the financial support by the Ontario Ministry of Environment and Energy, a Special University Scholarship by the University of Western Ontario, and the collaboration grant by Trent University.

Finally, I would like to thank my family for their love and encouragement. My deepest appreciation goes to my parents who have always inspired me to pursue my career. Last but far from least, I thank my wife Renee and my son Mark, for their love, belief and solid support that helped me through the difficult times, without them I could not have gone this far now.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACexpert	Analytical Chemistry expert *system shell*
ACmethod	Analytical Chemical method expert *system module*
AGC	Automatic Gain Control
API	Application Programming Interface
BBISO	Broad Band ISOlation *waveform*
CAD	Collisionally Activated Dissociation
CCPA	*The* Canadian Chemical Producer's Association
CE%	Conversion Efficiency (%)
d.c.	direct current
DDE	Dynamic Data Exchange
DIAGplatform	DIAGnostic platform *modular program*
DLL	Dynamic Linking Library
DOS	Disk Operating System
EAengine	Environmental Analytical *inference* engine
EAshell	Environmental Analytical shell
ECOexpert	ECOlogical expert *system module*
EI	Electron *impact* Ionization
ERAPs	Emergency Response Assistance Plans
ERexpert	Emergency Response expert *system module*
FID	Flame Ionization Detector
GC	Gas Chromatography
GC-ITMSD	Gas Chromatography-Ion Trap Mass Spectrometer Detector
GC-MS	Gas Chromatography-Mass Spectrometry
GCdiagnosis	Gas Chromatography diagnosis *expert system*
GCMSexpert	Gas Chromatography-Mass Spectrometry expert *system shell*
GLP	Good Laboratory Practice
GUI	Graphical User Interface
ITMS	Ion Trap Mass Spectrometer

KBF	Knowledge Base File
KDM	Knowledge Domain Matrix
KDS	Knowledge Delivery System
LIMS	Laboratory Information Management System
m/z ratio	mass *to* charge ratio
MSD	Mass Spectrometer Detector
MS^n, MSMS	Tandem Mass Spectrometry ($n \geq 2$)
OLE	Object Linking *and* Embedding
QISMS	Quadrupole Ion Storage Mass Spectrometer
QISMSdiagnosis	Quadrupole Ion Storage Mass Spectrometry diagnositic *expert system module*
r.f.	radio frequency
REMexpert	REMedial expert *system module*
RHCs	Representative Hazardous Chemicals
SPECview	SPECtra view *modular program*
SPI	*Varian* Septum Programable Injector
SPILLexpert	SPILL *response* expert *system shell*
TCDD	Tetra-Chloro-Dibenzo-Dioxins
TCDF	Tetra-Chloro-Dibenzo-Difurans
TIC	Total Ion Chromatogram
t_R	retention time

CHAPTER 1
INTRODUCTION

1.0 GENERAL INTRODUCTION

Artificial intelligence is the application of mathematical and logical techniques by man-made systems to carry out 'reasoning' abilities similar to those of human beings. This includes the ability to recognize objects, phenomena and situations, to analyze relationships between them, and to make decisions, even to predict the consequences of such decisions, and to disclose new regularities between observations. As shown in Figure 1.1, the major tasks of artificial intelligence can be divided primarily into the following four groups [1]:

1) Pattern recognition;
2) Machine adaptation and learning;
3) Problem solving;
4) Natural language understanding and verbal communication.

Expert system technology, which is a rich subset of artificial intelligence, is related to the task of problem solving. Expert systems are simply computer programs that are designed to emulate human experts through the integration of the knowledge of domain experts about a rather narrow field of study to furnish such programs with highly specific problem solving abilities [2]. A characteristic difference between an expert system and a conventional computer program is the strict separation of the knowledge from the program code. This principle governs the basic architecture of an expert system, no matter how large or how complex the system, an expert system always contains three fundamental parts: a knowledge base, an inference engine, and a user interface. As shown in Figure 1.2, the knowledge base of an expert system contains both factual information and heuristic knowledge about a problem domain. In some circumstances, the expert system may need to supplement the heuristic knowledge in the knowledge base by accessing external databases. The

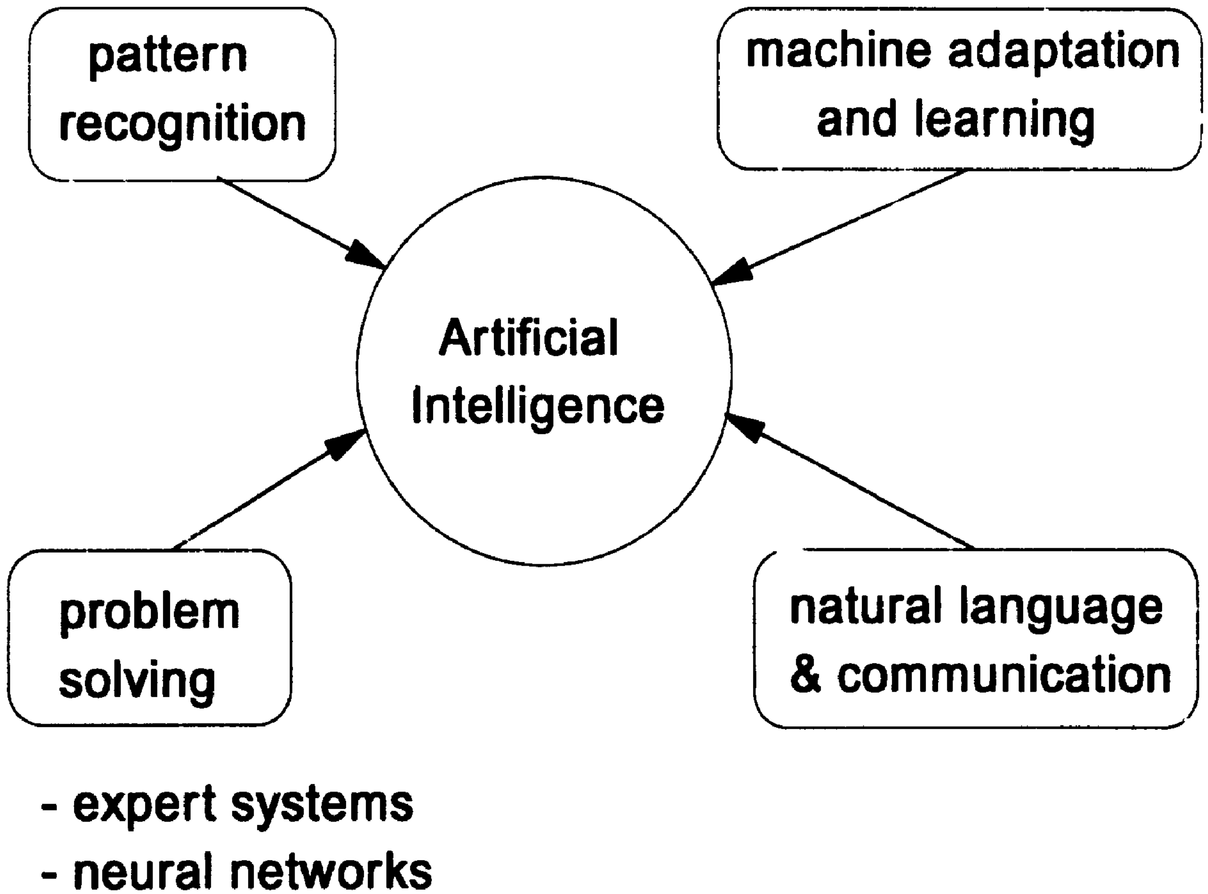


Figure 1.1 A description of the primary tasks involved in application of artificial intelligence (AI)
Expert system technology, as a rich subset of AI, is related to the task of problem solving.

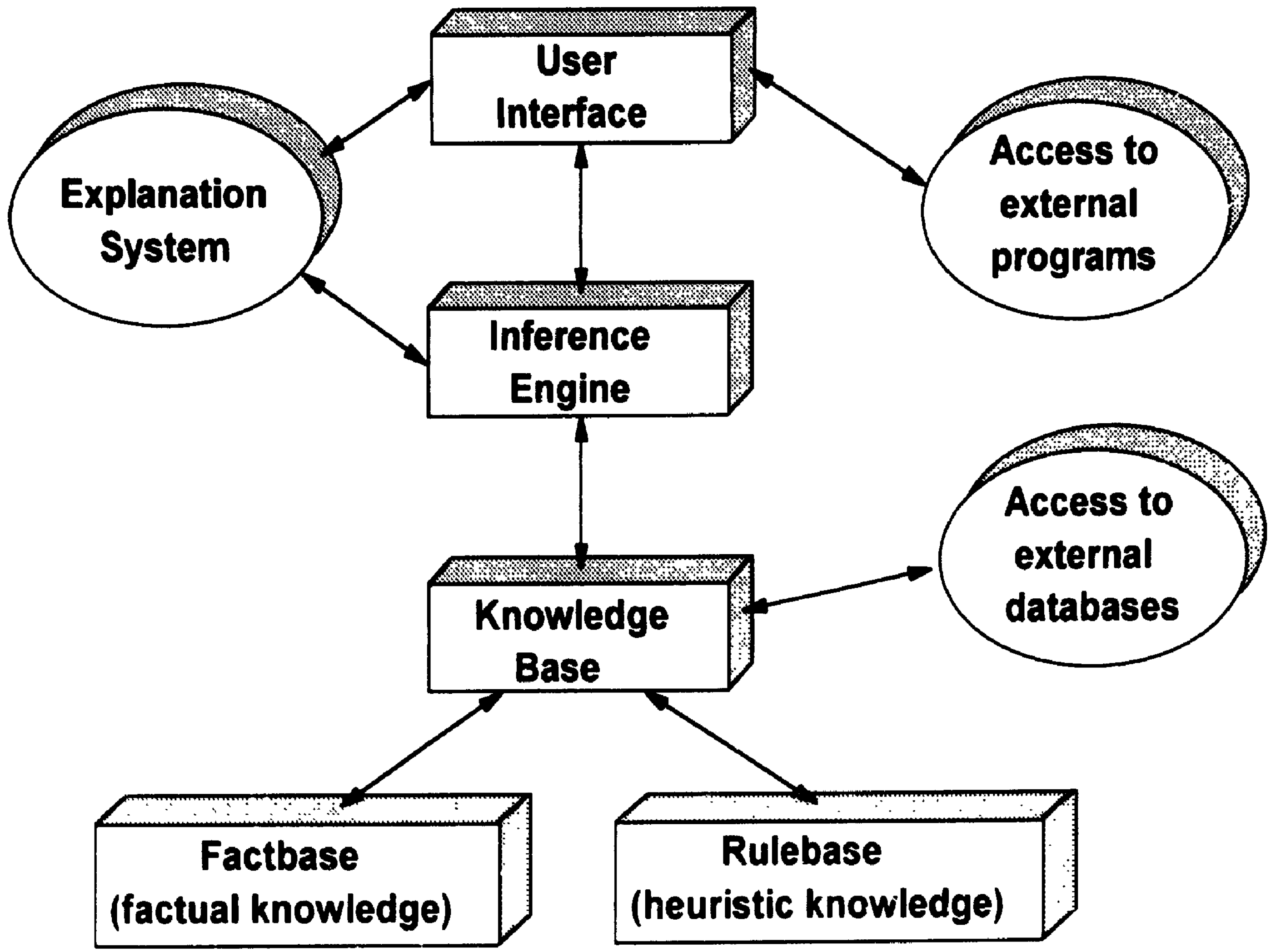


Figure 1.2 The principal components in an expert system.

inference engine is the reasoning component of an expert system that uses the knowledge in the knowledge base (often in a structured format) to draw conclusions from a set of conclusions in ambiguous situations. The user interface provides system users with an efficient operational environment.

1.1 SOFTWARE FOR THE DEVELOPMENT OF EXPERT SYSTEMS

While it is possible to construct an expert system using a procedural language, such as FORTRAN or C, it is difficult to encode heuristic human knowledge using these conventional languages. The principle of a strict separation of knowledge and control code also determines the nature and structure of the tools that can be used in the development of expert systems. Software tools may be classed into three groups: i) procedural languages, ii) higher-level programming languages, and iii) expert system shells.

The nature of expert system development requires software tools that can be more efficient at handling data in a symbolic format, rather than at handling numeric calculations in a procedural manner. Theoretically, any algorithm can be programmed using a conventional programming language, nevertheless the use of procedural languages for the development of an expert system is rather inefficient and more or less inappropriate. At a somewhat higher level are the languages that have been developed specifically for effective manipulation of symbolic data. These languages have a rich internal structure, such as lists, chains, sets, trees, mechanisms for returning during state-space searches, associative memory, as well as a range of inference mechanisms. In this category, languages like LISP, PROLOG, SMALLTALK, and OPS among others, are the most commonly used.

More recently, the development of universal modular systems, called expert system shells, has made the construction of an expert system much easier [3]. In such shells, some components are prefabricated and fixed without being firmly linked into the functional unit. An expert system shell contains an inference engine, an interface, and knowledge acquisition aids, but lacks rules and facts in the knowledge base [4]. In this case, the process of development is centered on the creation and modification of the knowledge base, and on linking each of the components into a functional unit. Detailed reviews in this regard have been reported by Settle and Van Leeuwan in which they discussed the expert system development tools available to chemists [5,6].

1.2 EXPERT SYSTEM APPLICATIONS IN THE CHEMISTRY DOMAIN

Since the first development of DENDRAL in the late 1960's, expert system technology has been increasingly incorporated into many different applications. In the chemistry domain, such applications span a broad spectrum, ranging from structure elucidation to structure and activity correlation, to data interpretation, synthesis planning, and experiment design. A detailed summary is given in Table 1.1.

1.2.1 Structure elucidation and data interpretation

One of the earliest examples of heuristic programming in chemistry was the DENDRAL project, which is now considered as a milestone in the history of expert system applications. DENDRAL was originally developed to assist structural organic chemists elucidate likely structures based on mass spectral data. Presently, DENDRAL is one of the most widely used expert systems that has helped in thousands of chemical structure analyses [1]. In fact, structure elucidation and data interpretation have been very active research areas in which application of expert system technology covers various spectral and chromatographic methods or combination of both and results in the generation of powerful new computer-based tools. The research activities in this area that have been reported during the last five years are briefly summarized in Table 1.1 [7-35].

Table 1.1. Applications of expert system technology in the chemistry domain

Sub Area	Topics	Examples
Structure elucidation and data interpretation	MS,[7-13, 27] IR,[14-18, 28] NMR,[19-23] UV,[24] Raman, [25] X-ray,[26] GC-MS,[29, 30], MS-MS,[32] Chromatography.[33-35]	DENDRAL,[9, 10, 27] MYCIN,[13] Meta-DENDRAL,[11, 12] HIPS,[21] MAXMASS,[7, 8] EXPIRS,[28] PEPTO,[23] AUTARG,[29] ESSESA,[20] Cathie,[35] PAIRS,[15,18] ESESOC,[31] RASTR.[24]
Computer assisted syntheses	Synthesis, retrosynthesis, starting material selection.[36-44]	LHASA,[45-47] SYNLIB,[48] SYNGEN,[49] FOWARD,[49] LILITH.[50, 161]
Analytical method development	HPLC,[51-70] IC,[71] GC, Electrochemical analysis,[72-77] UV,[77] IR, Ultracentrifugation,[79] Sampling[80] and Analysis.	CRISEbo..,[52] ECAT,[63] ESCA,[62, 64-68] RES,[69-70] OPSAES,[80] SpinPro,[79] ChromDream.[81]
Structure and activity correlation	Bioactivity,[82-86, 88-89] chromatographic[90-93] behavior vs. structure.	CRIPES,[92] META,[84-85] CASE,[82] HPLC-METABOL EXPERT,[91] HAZARDexpert.[86]
Knowledge-based chemical database	Encapsulate standards methods, sample preparation,[95-96] interactive retrieval information.[48,94,97-100]	FATE,[97] SYNLIB,[48] SCANNET.[98]
Experiment design	Procedure design,[101,103-104] method selecti. n.[102]	DXPERT[102]
Analytical process automation	Automation,[25,105-113] quality contro!,[14] process control,[123-126] data comparison,[115] problem diagnosis,[116-117] data acquisition.[127]	CAFD,[118-119] AAexpert,[111] ACexpert,[120] AAd..gnosis,[121] GCdiagnosis,[122] PAWMI,[113] AUTOMAT,[129] LABGEN,[114] H-FLO,[128] QISMSdiagnosis[130]
Tutorial expert systems	Interactive, graphical training and teaching systems.[131-133]	DENDRAL,[9, 10] LHASA,[45-47] IRexpert.[134]
Other applications	Physicochemical property,[135] identification,[136] naming,[137] chemical spill reponse.[138-139]	SOL,[135] ERexpert.[138-139]

1.2.2 Computer assisted synthesis

Computer Assisted Organic Synthesis (CAOS) has become an accepted tool in the design of synthetic pathways. Many different approaches are in use, however, they essentially follow the skeletal dissection route in which a target structure is linked to the functionality and reactivity through either synthetic or retrosynthetic reasoning based mainly on the research efforts of Corey and his group [36-40]. LHASA is one such expert system that uses the retrosynthetic route to direct syntheses towards the selection of particular starting materials [45-47]. LHASA is highly interactive allowing unlimited user input and assistance in the decision-making process as syntheses proceed. The program relies on a database of known reactions (about 1100 for LHASA-11) and involves either transform-oriented or structure-oriented strategies to direct the user to simplify molecular structures in terms of size, functionality, internal connectivity, stereorelationships, and chiral centers to reach simpler precursor structures [37,39].

1.2.3 Analytical method development

Within analytical chemistry, chromatography is a particularly active area where research is being undertaken to implement expert system technology. Many applications have been described that provide computer assisted analytical method development [51-81]. The ESCA project (Expert Systems for Chemical Analysis) is supported by ESPRIT (The European Strategic Program for Research and Development in Information Technology), which involves a group of experts in a joint international effort to demonstrate the applicability of expert systems in HPLC, and has resulted in the development of a number of sub-system modules [62,64-70]. The project is considered to have been successful and has revealed many useful conclusions that are important for the future development of expert system applications [64]. Among them is the concept of finding narrow focuses for development of modular systems and the systematic integration of these modules into a global framework. Another example in this area is the expert system developed for voltammetric determination of trace metals by Esteban and co-workers [72-77]. This system is built using the KES expert system shell to perform the method development task for both quantitative and qualitative electrochemical analysis of a wide range of metals.

1.2.4 Structure and activity correlation

The Quantitative Structure-Activity Relationship (QSAR) program implements expert system technology together with molecular and structural theories to predict trends of the target compounds in either biological activities [82-89], or chromatographic behavior [35,90-93]. Klopman, *et al.* have studied relationships between substructures and bioactivities [82,87]. The META program was developed to predict the sites of potential enzymatic attack and the nature of the chemicals formed by such metabolic transformations [84-85]. The expert system CRIPES was developed to predict retention time properties of organic molecules in reversed-phase HPLC using indices based on an alkyl-aryl-ketone scale derived from empirical quadratic expressions [92].

1.2.5 Knowledge-based chemical database systems

Traditional ways of using database systems only for data and information storage and retrieval can no longer satisfy the growing demand from real world situations Some have argued about the importance of developing a new generation of database systems that could fully benefit from the power of present day computational techniques [94]. These new systems would not only manage data and information more efficiently, but would also implement mathematical, logical, and chemical theories to encapsulate human knowledge, and, therefore, be able to extrapolate from the stored data and information to practical situations. Over the last five years or so, a number of applications have appeared [48, 95-100] SYNLIB is an interactive database system that allows complex chemical synthesis concepts to be presented directly to the user, and requires minimal computer expertise on the part of chemist [48]. Settle *et al.*

have developed a database of validated microwave sample dissolution methods that can be transferred electronically among different laboratories and is capable of reproducing procedures for specific sample types [100]. FATE is an on-line database system for interactive retrieval of kinetic and equilibrium constants that describe the fate of chemicals in the environment, that has been developed by incorporation of fundamental chemical structural theories into the expert system SPARC. The resulting program is claimed to be able to provide reliable and environmentally realistic fate constants that have not been available before [97].

1.2.6 Experiment and procedure design

Deciding on the selection of optimal laboratory procedures is not a simple task even for an experienced analyst accounting the large number of parameters that must be considered [101]. Expert system technology has been implemented in this area to capture the heuristic knowledge applied by the expert. For example, DXPERT, which uses fuzzy logic concepts, has been developed to assist chemists rank thirteen experimental designs according to their feasibility for a given project [102]. Van den Boagert, *et al.* have developed an expert system that assists in the design of experiments using x-ray fluorescence techniques [103].

1.2.7 Analytical process automation

Unattended analysis is one of the ultimate goals of laboratory automation projects [33]. Before an instrument can carry out an automated analysis, the controlling system must be able to perform successfully a series of tasks while left unattended. The "successful" part requires that the control software can assess the quality of the results of on-going tasks, which include calibration of the instrument, modification of run-time parameters, analysis of feed-back signals, diagnosis of operational errors, and modification of sample introduction procedures. Various applications have been developed to address such tasks (see Table 1.1 for details). The Analytical Director program has been developed to automate the conversion of 'textbook' analytical procedures by applying chemical knowledge in a deep-reasoning-based inference mechanism [105-107]. PAWMI is a program developed to carry out automated identification of bulk organic waste mixtures using IR data as a preliminary screening tool prior to quantitative analysis [113]. A number of instrumental diagnostic expert systems have been developed. For instance, Tsuiji *et al.* have developed a program for troubleshooting HPLC assay methods [116]. The AAexpert program has been designed to implement totally automated analysis by flame AAS [111]. Over the last ten years, our laboratory has developed a number of expert systems serving diagnostic purposes for different types of analytical instrumentation including GC, GC-MS, and flame AAS [120-122,130]. For a more detailed description, see Section 1.3.

1.2.8 Tutorial expert systems and other applications

Possibly the most significant advantage of developing an expert system is that expertise from one or more domain experts can be effectively captured and encoded. With the help of graphical functionality, tasks like teaching, simulation, and on-line tutorials can be programmed in a more efficient and direct manner. Programs like DENDRAL [9-10], LHASA [45-47], and IRexpert [134], can be used as interactive tutoring systems to train novice laboratory personnel and students.

Expert systems have also been applied to a number of other areas. Applications that allow evaluation of physico-chemical properties [135], characterization of polymers [136], converting chemical formulae to names [137], and handling emergency chemical spill accidents [138] have been described. Some standard procedures in the chemical decision-making stream can be taken over by expert systems, improving the performance and productivity in laboratories, as well as increasing the extent of knowledge sharing. The overall success of current applications of expert system technology suggests that this technology will be integrated more frequently into the software support used in chemistry and analytical laboratories.

1.3 APPLICATIONS IN ANALYTICAL CHEMISTRY-THE ACexpert PROJECT

For nearly twenty years, our laboratory has been involved in the application of computer technology in instrumental analysis [111,120-122,130,138-148]. The early work, such as CDSCAN, SpectraManager, and Simpfit were computer programs developed specifically for instrument control and manipulation of spectral data using the concept of spreadsheets [141-144]. These programs dealt with problems that are more procedurally and numerically oriented and common in experimental spectroscopic research. A large number of experimental papers have made use of these programs to report spectroscopic data [149-150]. In 1987 we began to encounter problems that were more symbolic in nature as the ACexpert project was developed [120]. We used initially the commercially available expert system shell KDS3 (KDS Corp., 934 Hunter Rd., Wilmette, IL 60091). However, the lack of a graphic user interface made development difficult [121,139]. Therefore, in 1991 we designed an expert system shell that could fit the needs of development of chemical applications for use based on IBM PC or compatible machines running the Microsoft Windows operating environment. EAshell was developed using Microsoft C (version 7.0) and has been central to our work since 1991.

1.3.1 A description of the ACexpert project

Automation of analytical processes calls for real time communication between the instrument and the control program. The integration of expert systems into automated systems will become increasingly commonplace. The goals of the ACexpert project were to study the feasibility of implementing expert system technology in various stages of the chemical analysis process. The development began with an investigation of a generic analytical process. We identified the following stages involved in chemical analysis:

(1) Sampling;

(2) Separation and preparation;

(3) Selection of analytical method(s);

(4) Performing the analysis;

(5) Data analysis for QA/QC;

(6) Reporting analytical results and predicting future trends if applicable;

(7) Delivering training for laboratory personnel (optional).

According to this definition, the preliminary components of the overall controlling expert system were specified and assembled in a global structure named ACexpert [120,144,145]. Figure 1.3 describes the ACexpert structure in which the relationships between the individual components are illustrated. Within this structure, each modular expert system performs a separate task in different analytical stages and is capable of acting independently or as a module of the full system A summary of the roles of the expert system modules and their functions is shown in Table 1.2.

Table 1.2. Description of specific roles of expert system used in ACexpert and the individual tasks to be performed in an analytical laboratory

Role	Module	Use
Consultation	**ACselect, ACmethod**	Selecting sample preparation and analysis methods from a list of ingredients or procedures.
Control	**ACcontrol**	Directing analytical process control.
	ACdiagnosis	Diagnosing probable causes of malfunctions or incorrect data values.
Analysis	**ACanalysis, AC-QC**	Collecting analytical results, evaluating data quality/extracting information, inferring consequences from a given situation.
Report	**ACtrend, ACreport, ACassurance**	Reporting evaluated analytical data and information, appling guide line to assure process/production qualty.
Instruction	**ACteach**	Training laboratory personnel or teaching students through simulations and Q&A sessions.

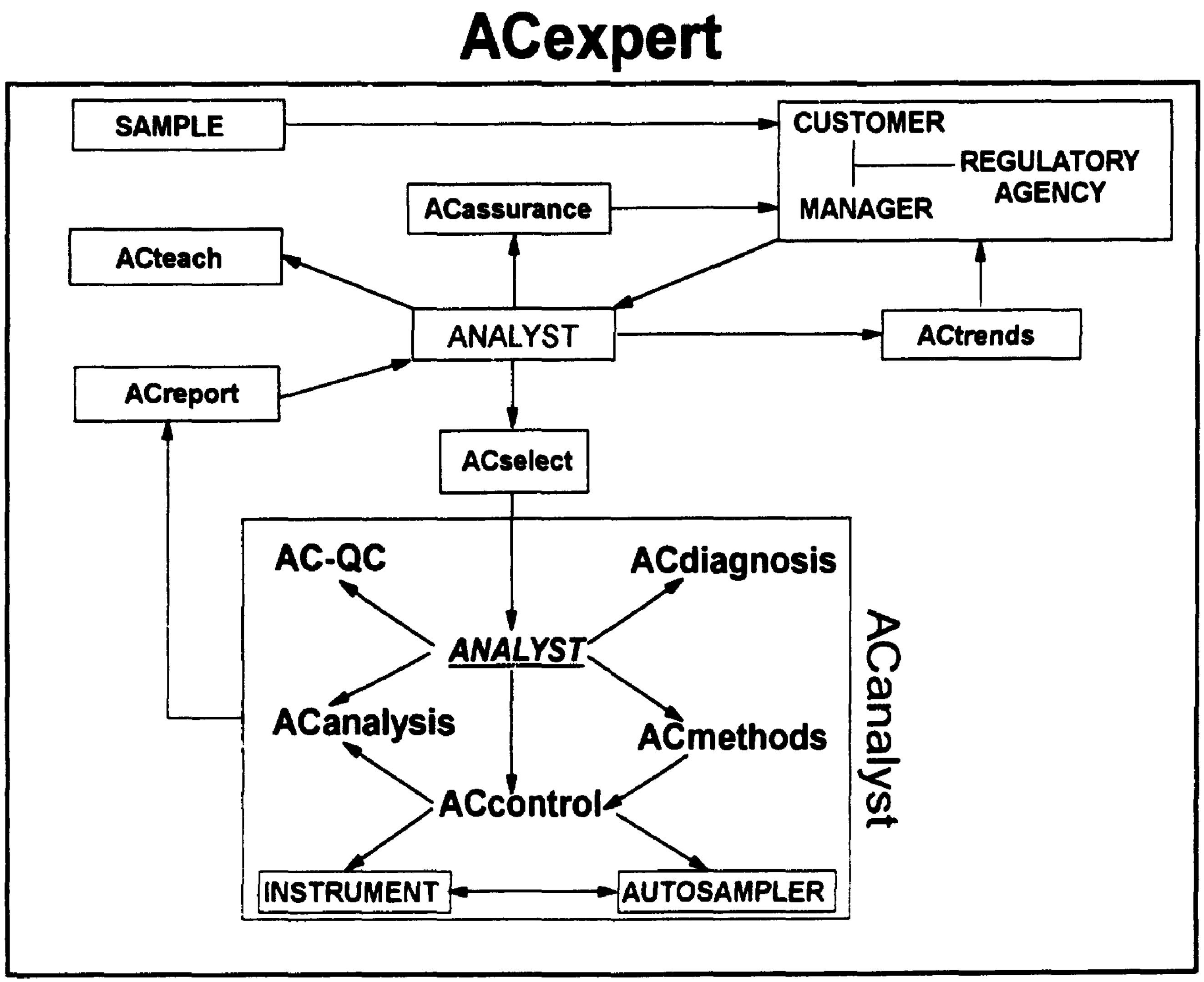


Figure 1.3 The structure of the ACexpert project.

The principal tasks identified in a generic analytical process are arranged into this structure. Corresponding modular expert systems are designed in this global program, some of them are topics of this thesis.

Figure 1.4 shows the development philosophy that implements (1) the KDM scheme for the knowledge acquisition and representation and (2) a graphical user interface to access internal and external databases and that also provides communication with users for input and output purposes.

1.3.2 Development of a Windows-based expert system shell

The KDM knowledge coding scheme was first implemented in the development of a Microsoft Windows (3.1) based expert system shell, named EAshell [140]. Microsoft C version 7.0 was used for the development. The shell currently comprises three components, namely EAengine, TableGenerator, and RuleEditor. EAengine is an inference engine that contains a library of subroutines that functions as a dynamic linked library (DLL) within Windows. EAengine supports the three inference strategies, namely, forward chaining, backward chaining, and mixed chaining. According to the domain problem, these inference strategies can be specifically selected or used interchangeably. The TableGenerator provides a blank tabular matrix for the knowledge engineer to input domain knowledge following the KDM concept outlined above [140]. RuleEditor is used to convert automatically the encoded knowledge in a KDM into a rulebase of production rules, a KBF.

1.3.3 Summary of applications developed under the ACexpert project

Using the scheme shown in Figure 1.4 with the inference engine provided by EAshell, a number of Windows-based expert system prototypes have been developed in our laboratory to address task-specific problems in different analytical processes, these include AAexpert [111,120-121,144], GCdiagnosis [122,146], SPILLexpert [138-139], and GCMSdiagnosis [130]. Table 1.3 is a summary that specifies the names of applications developed, the software tools used, and the analytical tasks each prototype system addressed, as well as the key personnel involved in the development.

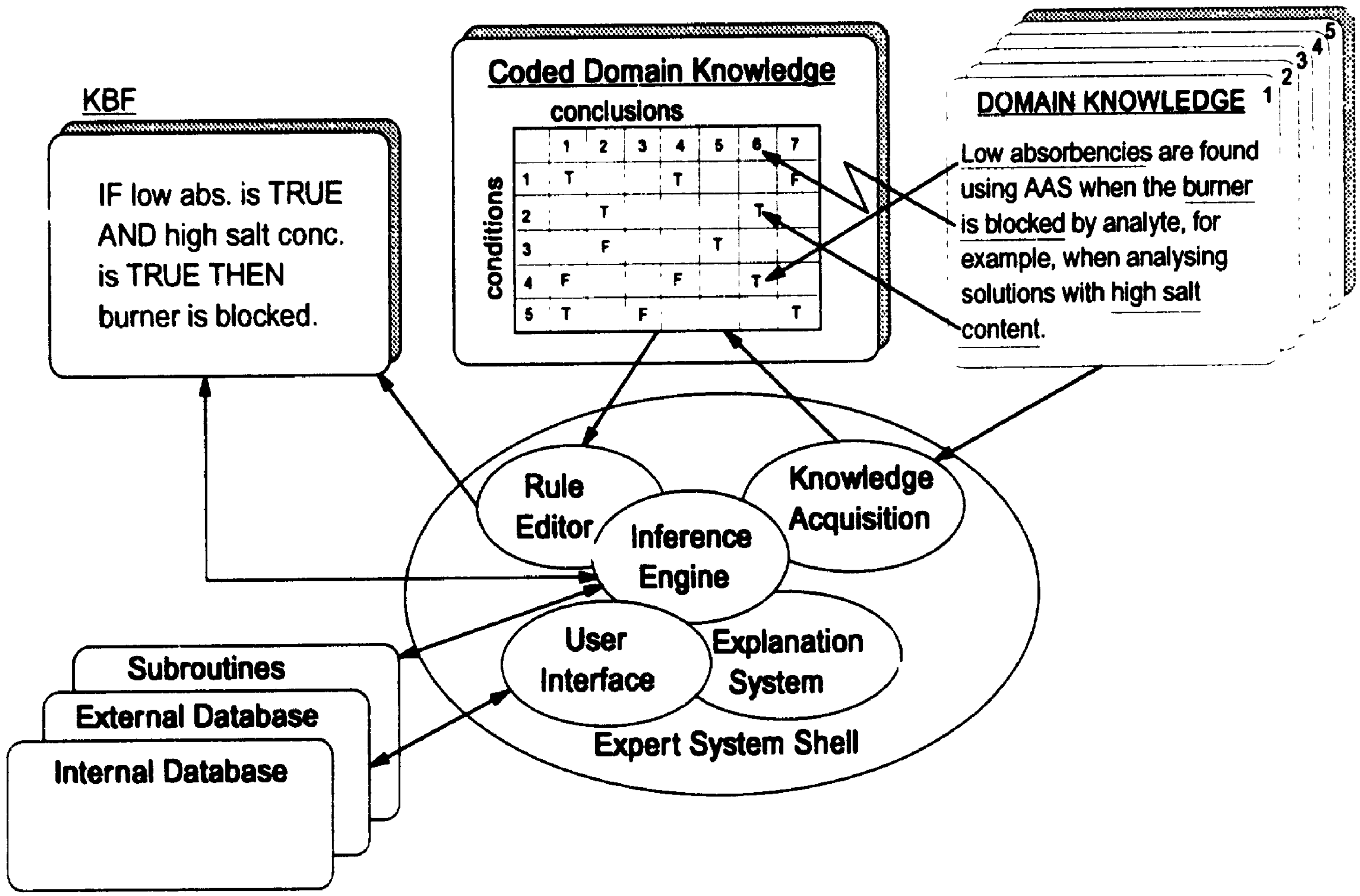


Figure 1.4 The philosophy implemented in the development of expert system applications under the ACexpert global structure using the EAshell components.

Table 1.3. A summary of the applications developed in the our laboratory over the past six years.

Program (personnel)	**Module**	**Software Tool**	**Description**
EAshell (G. Huang)	EAengine (140)	C	Windows-based inference engine, accessible by DLL function call.
	TableGenerator (147,148)	C, Excel	A tool kit for the knowledge acquisition process.
	RuleEditor (147,148)	C	A batch processor to convert a filled KDM into a rule file (KBF).
AAexpert (W. Browett, S. Lahiri)	AAdiagnosis (120,121)	KDS, EAshell, Visual Basic	A diagnostic expert system for the atomic absorption spectrometer.
	AAmethod	EAshell, Visual Basic	An expert system for method selection in flame AAS.
	AAcontrol (111,141,142)	EAshell, C, FORTRAN	A control program for automated AAS analysis of trace metals.
GCexpert (H. Du)	GCdiagnosis (122,146)	EAshell, C, Visual Basic	A diagnostic expert system for gas chromatography.
	GC-QC	EAshell, C	A module for GC data analysis.
SPILLexpert (Q. Zhu)	ACselect (145)	Quick Basic, KDS	A module for analytical methods selection based on matrix, concentration, and detection limit.
	ERexpert (138,139)	EAshell, Access, Visual Basic	A program uses an internal data-base and expert system module to advise on the best response to emergency chemical spills.
	ACmethod (145)	EAshell, KDS, Access, Visual Basic	A knowledge system comprising a database of methods and an expert system component.
GCMSdiagnosis (Q. Zhu)	QISMSdiagnosis (130)	EAshell, Visual Basic	A diagnostic module for ion-trap MSMS analysis.
	SPECview (130)	Visual Basic	A module to access directly Varian GCMS data.
	DIAGplatform (130)	Visual Basic	A module to calculate conversion efficiency of GC-MS-MS process.
	RuleChecker	EAshell, Visual Basic	A Window program to detect in-consistencies in a knowledge file.

We have found that the use of EAshell components has allowed the successful development of a series of modular systems that address different analytical problems involving both consultation and control. AAmethod and ACselect are modules performing method selection tasks for different analytical problems [145]. AAdiagnosis, GCdiagnosis and QISMSdiagnosis are programs that diagnose the causes of instrumental malfunction [121-122,130]. AAcontrol automatically controls an AAS analysis process [111,120-121,144]. ERexpert and ACmethod advance our applications by combining expert system technology with database systems to solve specific chemical problems [138,139]. Presently, we are focusing on tasks related to analytical data analysis, data quality evaluation, and trend prediction. The SPECview module provides direct data accessibility for the Varian Saturn series of GC-MS instruments [130]. The DIAGplatform module calculates the conversion efficiency of an MS-MS process and can be used as an indicator to provide the operational status of the GC-MS instrument. Both programs can be connected to the QISMSdiagnosis module when a low conversion efficiency is detected. The diagnostic program can be activated to diagnose the possible cause(s) of the problem(s) for the user to rectify. GC-QC is a module under development that aims at evaluating the quality of gas chromatographic data.

1.4 SCOPE OF THESIS

The work in this thesis is aimed at understanding the process of encoding chemical knowledge into computer readable forms and subsequently developing knowledge-based applications to solve domain-specific problems in sub-areas in the analytical and environmental chemistry field. The thesis as a whole can be divided into two parts. Part one (Chapter 2 & 3) focuses on the methodology of our research, and part two (Chapter 4 & 5) describes the implementation of these methods into applications.

In Chapter 2, the process of the transfer of human knowledge into computer data structure is examined. Our results reveal that three common steps are generally involved, (a) defining and understanding a problem subdomain (the target), (b) acquisition of the knowledge, and (c) representation of the acquired knowledge in a specific logical format that a computer can understand. Based on this understanding, an innovative knowledge encoding scheme, named the knowledge domain matrix (KDM), has been developed and implemented. The importance of a user interface and its role in the development of an expert system is addressed in Chapter 3. Chapter 4 reports the development of an expert system in the environmental area. ERexpert is a knowledge-based program for use in response to emergency chemical spills. Next, the development of a diagnostic expert system for the GC-MS instrument is presented in Chapter 5. Finally, in Chapter 6, the significance of this work is summarized and the future direction of the application and research of expert system technology in the chemistry domain is discussed. To my knowledge and understanding, I believe that expert system technology will be more fully integrated into the chemistry domain in the future, with applications in analytical chemistry leading the way.

1.5 REFERENCES

1. Jakus, V. *Collect. Czech. Chem. Commun.*, **1992**, 57, 2413-2451
2. Penmetsa, R. K.; Grenney, W. J. *J. Environ. Eng.-ASCE*, **1993**, 119, 231-247.
3. Hohne, B. A.; Pierce, T. H. In *Expert System Applications in Chemistry*, Hohne, B. A.; Pierce, T. H., Eds.; American Chemical Society Washington, DC, 1989; ACS Symposium Series 408, Chapter 1, p 1.
4. Van Leeuwan, J. A.; Buydens, L. M. C.; Vandeginste, B. G. H.; Kateman, G. *Trends Anal. Chem.*, **1990**, 9, 49-54.
5. Settle, F. A. Jr.; Pleva, M. A. *Chemomtr. Intell. Lab. Syst.*, **1991**, 11, 13-26
6. Van Leeuwan, J. A.; Vandeginste, B. G. M.; Postma, G. J.; Kateman, G *Chemomtr. Intell. Lab. Syst.*, **1989**, 6, 239-252.
7. Scott, D. R. *Chemomtr. Intell. Lab. Syst.*, **1991**, 12, 189-200.
8. Scott, D. R. *Anal. Chim. Acta*, **1991**, 246, 391-403.
9. Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. *Application of Artificial Intelligence for Organic Chemistry: The DENDRAL Project.* McGraw-Hill, New York, 1980.
10. Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J *Artif Intell.*, **1993**, 61, 209-261.
11. Gray, N. B. A. *Chemomtr. Intell. Lab. Syst.*, **1988**, 5, 11-32
12. Feigenbaum, E. A.; Buchanan, B. G. *Artificial Intell.*, **1993**, 59, 233-240
13. Buchanan, B. G.; Shortliffe, E. H. *Rule Based Expert Systems. The Mycin Experiments of the Standford Heuristic Programming Project.* Addison-Wesley, Reading, MA, 1984.
14. Pyo, D. *Vibr. Spectrosc.*, **1993**, 5, 263-273.
15. Ying, L. S.; Levine, S. P.; Tomellini, S. A.; Lowry, S. R. *Anal. Chem.*, **1987**, 59, 2197-2203.
16. Moore, D.; White, J. S.; Harbin, B. A. *Anal. Chim. Acta*, **1994**, 294, 85-94
17. Hong, H. X.; Xin, X. Q. *J. Chem. Inf. Comput. Sci.*, **1990**, 30, 203-210

18. Tomellini, S. A.; Stevenson, J. M.; Woodruff, H. B. *Anal. Chem.*, **1984**, 56, 67-70.

19. Wehrens, R.; Buydens, L.; Kateman, G. *Chemomtr. Intell. Lab. Syst.*, **1991**, 12, 57-67.

20. Hong, H. X.; Xin, X. Q. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 1259-1266.

21. Wehrens, R.; Lucasius, C.; Buydens, L.; Kateman, G. *Anal. Chim. Acta*, **1993**, 277, 313-324.

22. Zupan, J.; Buydens, L.; Peeters, A.; Massart, D. L. *Chemomtr. Intell. Lab. Syst.*, **1990**, 8, 43-51.

23. Catasti, P.; Carrara, E.; Nicolini, C. *J. Comput. Chem.*, **1990**, 11, 805-818.

24. Elyashberg, M. E.; Serov, V. V.; Martirosyan, E. R.; Zlatina, L. A.; Karasev, Yu. Z.; Koldashov, V. N.; Yampolskiy, Yu. Yu. *J. Mol. Struct. (Themchem)*, **1991**, 230, 191-203.

25. Luinge, H. L. *Vibr. Spectrosc.*, **1990**, 1, 3-18.

26. Alder, B.; Schutze, P.; Will, J. *Anal. Chim Acta*, **1993**, 271, 287-291.

27. Lederberg, J.; Sutherland, G. L.; Buchanan, B. G.; Feigenbaum, E. A.; Robertson, A. U.; Duffield, A. M.; Djerassi, C. *J. Am.Chem. Soc.*, 1969, 91, 2973-2976.

28. Andreev, G. N.; Argirov, O. K.; Penchev, P. N. *Anal. Chim. Acta*, **1993**, 284, 131-136.

29. Stan, H-J.; Heberer, Th. *J. Chromatogr.*, **1993**, 653, 55-62.

30. Holzer, G.; Bertsch, W.; Zhang, Q. W. *Anal. Chim. Acta*, **1992**, 259, 225-235.

31. Hu, C. Y.; Xu, L. *Anal. Chim. Acta*, **1994**, 298, 75-85.

32. Palmer, P. T.; Hart, K. J.; Enke, C. G.; Wade, A. P. *Talanta*, **1989**, 36, 107-116.

33. Elling, J. W.; Mnisezwski, S. M.; Zahrt, J. D.; Klatt, L. N. *J. Chromatogr. Sci.*, **1994**, 32, 213-218.

34. Scheuer, K. *Chemomtr. Intell. Lab. Syst.*, **1993**, 19, 201-216.

35. Milne, R. *J. Chromatogr.*, **1989**, 485, 341-347.

36. Corey, E. J. *Pure Appl. Chem.*, **1967**, 14, 19-37.

37. Corey, E. J.; Jeffery, Howe, W.; Orf, H. W.; Pensak, D. A., Petersson, G. *J. Am. Chem. Soc.*, **1975**, 97, 6116-6124.

38. Corey, E. J.; Wipke, W. T. *Science*, **1969**, 166, 178-192

39. Corey, E. J.; Long, A. K.; Rubenstein, S. D. *Science*, **1985**, 228, 408-418

40. Corey, E. J.; Cheng, X. M. *The Logic of Chemical Synthesis*, Wiley New York, 1989.

41. Ott, M. A.; Noordik, J. H. *J. Royal Netherlands Chem Soc.*, **1992**, 111, 239-246.

42. Dogane, I.; Takabatake, T.; Bersohn, M. *J. Royal Netherlands Chem Soc.*, **1992**, 111, 291-296.

43. Bador, P.; Surrel, M. N.; Lardy, J. P. *New J. Chem.*, **1992**, 16, 413-423

44. Pollet, P. *New J. Chem.*, **1990**, 14, 957-963.

45. Johnson, A. P.; Marshall, C.; Judson, P. N. *J. Royal Netherlands Chem Soc.*, **1992**, 111, 310-316.

46. Johnson, A. P.; Marshall, C.; Judson, P. N. *J. Chem Inf Comput Sci.*, **1992**, 32, 411-417.

47. Corey, E. J.; Long, A. K.; Lotto, G. I.; Rubenstein, S. D. *J. Royal Netherlands Chem. Soc.*, **1992**, 111, 304-309

48. Chodosh, D. F.; Hill, J.; Shpilsky, L.; Mendelson, W. L. *J. Royal Netherlands Chem. Soc.*, **1992**, 111, 247-254.

49. Hendrickson, J. B.; Parks, C. A. *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 209-215.

50. Baumer, L.; Compagnari, I.; Sala, G.; Sello, G. Erythronolide-B *J Royal Netherlands Chem. Soc.*, **1992**, 111 297-303.

51. Hamoir, T.; Bourguignon, B.; Massart, D. L.; Hindriks, H. *J Chromatogr.*, **1993**, 633, 43-56.

52 Bourguignon, B , Vankeeberghen, P ; Massart, D. L. *J. Chromatogr.*, **1992**, 592, 51-57

53 Szepesi, G., Valko, K. *J. Chromatogr.*, **1991**, 550, 87-100.

54 Schoenmakers, P J.; Bartha, A., Billiet, H. A. H. *J. Chromatogr.*, **1991**, 550, 425-447

55 Van Leeuwen, J. A.; Vandeginste, B. G. M.; Kateman, G.; Mulholland, M.; Cleland, A. *Anal. Chim. Acta*, **1990**, 228, 145-153.

56 Mulholland, M ; Dunand, N.; Cleland, A.; Van Leeuwen, J. A.; Vandeginste, B G M. *J. Chromatogr.*, **1989**, 485, 283-296.

57 Williams, S. S., Karnicky, J. F.: Excoffier, J. L.; Abbott, S. R. *J. Chromatogr.*, **1989**, 485, 267-281.

58 Desmet, M., Musch, G.; Peeters, A.; Buydens, L.; Massart, D. L. *J. Chromatogr.*, **1989**, 485, 237-253.

59 Schoenmakers, P J.; Dunand, N. *J Chromatogr.*, **1989**, 485, 219-236.

60 Conti, P.; Hamoir, T ; DeSmet, M.; Piryns, H.; van den Driessche, N ; Maris, F ; Hindiks, H.; Schoenmakers, P J , Massart, D. L. *Chemomtr. Intell Lab. Syst.*, **1991**, 11, 27-35.

61. Zhang, Y.; Zou, H.; Lu, P J. *Chromatogr.*, **1990**, 515, 13-26.

62. Wehrens, R.; Buydens, L.; Hindriks, H.; Maris, F. *Chemometr. Intell. Lab. Syst.*, **1994**, 25, 341-354.

63 Bach, R.; Karnicky, J.; Abbott, S. In *Artificial Intelligence Application in Chemistry*, Pierce, T H.; Hohne, B. A., Eds.; American Chemical Society: Washington, DC, 1986; ACS Symposium Series 306, Chapter 22, p278.

64 Buydens, L.; Schoenmakers, P.; Maris, F.; Hindriks, H. *Anal. Chim. Acta*, **1993**, 272, 41-51.

65 Mulholland, M.; Walker, N.; Maris, F.; Hindriks, H.; Buydens, L ; Blaffert, T.; Schoenmakers, P. J. *J. Chromatogr.*, **1991**, 550, 257-266.

66 Maris, F ; Hindriks, R., Vink, J.; Peeters, A., van den Driessche, N ; Massart, L *J. Chromatogr.*, **1990**, 506, 211-221.

67. Buydens, L. M. C.; Van Leeuwan, J. A.; Mulholland, M.; Vandeginste, B G G.; Kateman, G. *Trends Anal. Chem.*, **1990**, 9, 58-62.

68. Van Leuwan, J. A.; Buydens, L. M. C.; Vandeginste, B G. M., Kateman, G *Anal. Chim. Acta*, **1990**, 235, 27-40.

69. Van Leeuwan, J. A.; Buydens, L. M. C.; Vandeginste, B. G M , Kateman, G.; Schoenmakers, P. J.; Mulholland, M. *Chemometr. Intell Lab Syst* , **1991**, 11, 37-55.

70. Van Leeuwan, J. A.; Buydens, L. M. C.; Vandeginste, B G. M ; Kateman, G.; Cleland, A.; Mulholland, M.; Jasen, C.; Maris, F. A.; Hoogkamer, P H , van den Berg, J. H. M. *Chemometr. Intell. Lab. Syst.*, **1991**, 11, 161-174

71. Mulholland, M.; Haddad, P. R.; Hibbert, D. B. *J. Chromatogr* , **1992**, 602, 9-14.

72. Ruisanchez, I.; Larrechi, M. S.; Rius, F. X.; Esteban, M *Trends Anal Chem.*, **1992**, 111, 135-142.

73. Esteban, M.; Ruisanchez, I.; Larrechi, M. S.; Rius, F. X *Anal Chim Acta*, **1992**, 268, 95-105.

74. Esteban, M.; Ruisanchez, I.; Larrechi, M. S.; Rius, F. X. *Anal Chim Acta*, **1992**, 268, 107-114.

75. Esteban, M.; Arino, C.; Ruisanchez, I.; Larrechi, M. S , Rius, F X *Anal Chim. Acta*, **1993**, 284, 435-443.

76. Esteban, M.; Arino, C.; Ruisanchez, I.; Larrechi, M S., Rius, F X *Anal Chim. Acta*, **1994**, 285, 193-208.

77. Esteban, M.; Arino, C.; Ruisanchez, I.; Larrechi, M. S , Rius, F X *Anal Chim. Acta*, **1994**, 285, 377-389.

78. Buydens, L.; Detaevernier, M.; Tombeur, D.; Massart, D L *Chemometr Intell. Lab. Syst.*, **1986**, 1, 99-108.

79. Martz, P. R.; Heffron, M.; Griffith, O. In *Artificial Intelligence Application in Chemistry;* Pierce, T. H.; Hohne, B. A., Eds.; American Chemical Society Washington, DC, 1986; ACS Symposium Series 306, Chapter 23, p297

80. Cheng, C. P.; Liang, Y. C. *J. Chinese Chem. Soc.*, **1994**, 41, 89-95

81. Galushko, S. V.; Kamenchuk, A. A.; Pit, G. L. *J. Chromatogr.*, **1994**, 660, 47-59.

82 Rannug, U.; Sjogren, M.; Rannug, A.; Gillner, M.; Toftgard, R.; Gustafsson, J. A.; Rosenkranz, H.; Klopman, G. *Carcinogesis*, **1991**, 12, 2007-2015.

83 Nakadate, M.; Hayashi, M.; Sofuni, T.; Kamata, E.; Aida, Y.; Osada, T.; Ishibe, T.; Sakamura, Y.; Ishidate, M. *Environ. Health Perspect.*, **1991**, 96, 77-79.

84. Klopman, G.; Dimayugo, M.; Talafous, J. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 1320-1325.

85 Talafous, J.; Sayre, L. M.; Mieyal, J. J.; Klopman, G. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 1326-1333.

86. Smithing, M. P.; Darvas, F. In *Food Safety Assessment;* Finley, J. W.; Robinson, S. F.; Armstrong, D. J., Eds.; American Chemical Society: Washington, DC, 1992; ACS Symposium Series 484, Chapter 19, p 191.

87. Klopman, G. *J. Am. Chem. Soc.*, **1984**, 106, 7315-7321.

88. Judson, P. N. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 148-153.

89. Payne, M. P.; Walsh, P. T. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 154-161.

90. Dave, K.; Stobuugh, J. E.; Riley, C. M. *J. Pharma. Biomed. Anal.*, **1992**, 10, 49-60

91. Valko, K.; Szabo, G.; Rohricht, J.; Jemnitz, K.; Darvas, F. *J. Chromatogr.*, **1989**, 485, 349-363.

92. Smith, R. M.; Burr, C. M. *J. Chromatogr.*, 1989, **485**, 325-340.

93. Csokan, P. P.; Darvas, F.; Csizmadia, F.; Valko, K. *Magazine. Separat. Sci.*, **1994**, 12, 40-67.

94. Hicks, M. G. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 32-38.

95. Penninckx, W.; Smeyers-Verbeke, J.; Massart, D. L.; Spanjers, L. G. C. W.; Maris, F. A. *Chemometr. Intell. Lab. Syst.*, **1992**, 17, 193-200.

96 Penninckx, W.; Smeyers-Verbeke, J.; Massart, D. L.; Spanjers, L. G. C W.; Maris, F. A. *Anal. Chim. Acta*, **1993**, 282, 417-422.

97. Kollig, H. P.; Kitchens, B. E.; Hamrick, K. J. *J. Chem Inf Comput Sci*, **1993**, 33, 131-134.

98. Debska, B.; Hippe Z. S. *J. Mol. Struct.*, **1992**, 267, 261-268.

99. Lee, J. R.; Isenhour, T. L.; Marshall, J. C. *J. Chem Inf. Comput Sci*, **1992**, 32, 148-153.

100. Settle, F. A. Jr.; Walter, P. J.; Kingston, H. M.; Pleva, M A , Snider, T , Boute, W. *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 349-354

101. Wellmann, R.; Wunsch, G. A *Trends Anal. Chem.*, **1986**, 5, 138-141

102. Olivero, R. A.; Seshadi, S.; Deming, S. N. *Anal. Chim. Acta*, **1993**, 277, 441-453.

103. Van den Bogaert, B.; Morsink, J. B. W.; Smit, H. C. *Anal Chim Acta*, **1992**, 270, 115-130.

104. Van den Bogaert, B.; Morsink, J. B. W.; Smit, H. C *Anal Chim Acta*, **1992**, 270, 107-113.

105. Zhou, T.; Isenhour, T. L.; Zamfirbleyberg, M.; Marshall, J C *J Chem Inf Comput. Sci.*, **1992**, 32, 79-87.

106. Zhou, T.; Isenhour, T. L.; Marshall, J. C *J Chem. Inf Comput Sci*, **1993**, 33, 569-576.

107. Zhou, T.; Isenhour, T. L.; Marshall, J. C. *J. Chem Inf Comput Sci*, **1994**, 34, 558-569.

108. Palys, M. J.; Bos, M.; Vanderlinden, W. E. *Anal. Chim Acta*, **1991**, 248, 429-439.

109. Palys, M. J.; Bos, M.; Vanderlinden, W. E. *Anal. Chim Acta*, **1993**, 283, 811-829.

110. Palys, M. J.; Bos, M.; Vanderlinden, W E *Anal Chim Acta*, **1993**, 284, 107-118.

111. Lahiri, S.; Yuan, B.; Stillman, M. J. *Anal. Chem.*, **1994** 66, 2954-2963

112. Eckerttilotta, S. E.; Isenhour, T. L.; Marshall, J C *Anal Chim Acta*, **1991**, 254, 215-221.

113 Puskar, M. A.; Levine, S. P.; Lowry, S. R. *Anal. Chem.*, **1986**, 58, 1981-1989.

114. Klaessens, J.; Sanders, J.; Vandeginste, B.; Kateman, G. *Anal. Chim. Acta*, **1989**, 222, 19-34.

115. Danzer, K.; Wank, U.; Wienke, D. *Chemometr. Intell. Lab. Syst.*, **1991**, 12, 69-79.

116. Tsuji, K.; Jenkins, K. M. K. *J. Chromatogr.*, **1989**, 485, 297-309.

117. Brandt, J.; Hitzmann, B. *Anal. Chim. Acta*, **1994**, 291, 29-40.

118. Isermann, R.; Freyermuth, B. *J. Dynamic Syst. Measure. Control Transact. ASME*, **1991**, 113, 620-626.

119. Isermann, R.; Freyermuth, B. *J. Dynamic Syst. Measure. Control Transact. ASME*, **1991**, 113, 627-633.

120. Browett, W. R.; Cox, T. A.; Stillman, M. J. In *Expert System Applications in Chemistry;* Hohne B. A.; Pierce T. H., Eds.; American Chemical Society: Washimgton, DC, 1989; ACS Symposium Series 408, Chapter 17, p210.

121. Lahiri, S.; Stillman, M. J. *Anal. Chem.*, **1992**, 64, 283A-291A.

122. Du, H.; Lahiri, S.; Huang, G.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 21-31.

123. Peris, M.; Maquieira, A.; Puchades, R.; Chirivella, V.; Ors, R.; Serrano, J.; Bonastre, A. *Chemometr. Intell. Lab. Syst.*, **1993**, 21, 243-247.

124. Dewinter, M.; Grietens, D.; Rijckaert, M. *Trends Anal. Chem.*, **1990**, 9, 54-58.

125. Xu, L.; Xiao, Y. H.; Li, D. Q. *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 437-442.

126. Moors, M.; Massart, D. L. *Trends Anal. Chem.*, **1990**, 9, 164-169.

127 Koutny, L. B.; Yeung, E. S. *Anal. Chem.*, **1993**, 65, 148-152.

128. Pingand, P. B.; Lerner, D. A. H-FLO. *Anal. Chim. Acta*, **1994,** 293, 301-309.

129. Peris, M.; Chirivella, V.; Martinez, S.; Bonastre, A.; Ors, R.; Serrano, J. *Chemometr. Intell. Lab. Syst.*, **1994**, 26, 123-127.

130. Zhu, Q.; Du, H.; Plomely, J. B.; March, R. E.; Stillman, M. J. GCMSdiagnosis - An expert system for diagnosis of problem GC-MS data. *Anal. Chim. Acta*, **1996**, submitted.

131. Brockwell, J. C. *Abstract of Papers of the Am Chem. Soc.*, **1992**, 203, 32-CINF.

132. Collins, A. G.; Searleman, J.; Collins, K. J. *Water Sci. Tech.*, **1991**, 24, 307-314.

133. Sun, Y. D.; Haton, M. C.; Devin, B. *New J. Chem.*, **1990**, 14, 941-944.

134. Cabrol, D.; Rabine, J. P.; Richard, D.; Rouillard, M. *J. Chem Educat.*, **1993**, 70, 120-125.

135. Heller, S. R.; Bigwood, D. W.; May, W. E. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 627-636.

136. Ramsbottom, D. J.; Adams, M. J.; Carroll, J. *Trends Anal. Chem.*, **1994**, 13, 185-189.

137. Eggert, A. A.; Jacob, A. T.; Middlecamp, C. H. *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 227-233.

138. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 945-955

139. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 956-968.

140. Huang, G.; Stillman, M. J. An expert system shell under development in our laboratory, unpublished results.

141. Browett, W. R.; Stillman, M. J. *Comput. Chem.*, **1987**, 11, 73-82.

142. Browett, W. R.; Stillman, M. J. *Comput. Chem.*, **1987**, 11, 241-250.

143. Gasyna, Z.; Stillman, M. J. *J. Photochem.*, **1987**, 38, 83-98.

144. Browett, W. R.; Stillman, M. J. *Progr. Anal. Spectros.*, **1989**, 12, 73-110

145. Zhu, Q.; Stillman, M. J. Design and Implementation of ACselect In *Proceedings of the Technique Transfer Conference, The Ontario Minitry of Environment, Toronto, Ontario*, November 25-26, **1991**, 704-707.

146. Du, H.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 33-41

147. Du, H.; Huang, G.; Stillman, M. J *Anal. Chim Acta*, **1995**, Submitted

148. Lahari, S., Stillman, M. J. *Anal. Chim. Acta.*, **1995**, submitted.

149 Stillman, M. J.; Shaw, C. F., III; Suzuki, K. T. *Metallothioneins. Synthesis, Structure and Properties of Metallothioneins, Phytochelatins, and Metal-Thiolate Complexs.* VCH Publishers: New York, 1992.

150. Stillman, M. J. In *The Phthalocyanines: Properties and Applications, Vol. 3*; Leznoff, C. C.; Lever, A. B. P., Eds.; VCH Publishers: New York, 1993; Chapter 5, p230.

151. Sello, G. *J. Chem. Inf. Comput Sci.*, **1994**, 34, 120-129.

CHAPTER 2
KNOWLEDGE ACQUISITION: A KEY STEP IN EXPERT SYSTEM DEVELOPMENT AND THE KNOWLEDGE DOMAIN MATRIX SCHEME

2.0 INTRODUCTION

The development of DENDRAL thirty years ago, a system that assists in the interpretation of mass spectra data, is considered a landmark in the history of expert system applications [1,2]. Since then expert system technology has been increasingly incorporated into many different applications and is now an accepted vehicle for providing expertise on a range of domain problems. In the field of chemistry most of the current expert system applications deal with data interpretation, structure elucidation, synthesis planning, or instrument diagnosis, all areas in which the target problems are highly domain-specific [3-8]. However, there are problems, such as those found in the environmental chemistry domain that are less domain-specific and solving them requires multi-disciplinary knowledge

Expert systems are sophisticated computer programs that manipulate knowledge to solve problems efficiently and effectively in a narrow problem area [9]. The aim of an expert system is not merely to capture a static representation of a knowledge domain but to simulate the particular problem-solving task (or tasks) carried out within that domain [10]. Knowledge representation is the keystone of the enterprise of knowledge-based systems Some way of representing knowledge must be chosen for any program with a knowledge based content [11].

Knowledge acquisition is a crucial stage in the development of an expert system application [12]. As a process, it involves eliciting, analyzing, and interpreting the knowledge that a human expert uses when solving a domain problem and then transforming this knowledge into a suitable machine representation. Knowledge acquisition is critical since the ability and utility of a resulted expert system depends on the quality of the underlying representation of expert knowledge.

In order to have a better understanding of the knowledge acquisition process, this chapter first discusses briefly the knowledge representation methods, as well as inferential mechanisms used by expert system applications. The knowledge acquisition process will then be addressed with emphasis on the difficulties embedded in this process. Later in this chapter, a novel knowledge representation scheme developed at U.W.O., the knowledge domain matrix (KDM) approach, is described. Implementation of the KDM mechanism in the construction of expert systems are given in Chapters 4 and 5 of this thesis.

2.1 KNOWLEDGE REPRESENTATION METHODS

The challenge in the whole knowledge acquisition process lies in finding an effective and efficient knowledge representation scheme that allows not only smooth knowledge transfer (expert to computer), but also an effective use of the knowledge (computer to user). The choice of the knowledge representation scheme can have a substantial effect on the performance and utility of the resulting expert system. Generally, the methods used to represent knowledge may be divided into three categories, namely semantic networks, frame representations, and logic expressions using a rule format [9].

2.1.1 Semantic networks

The term semantic net is used to describe a knowledge representation method based on a network structure. A semantic network is defined as an oriented graph with evaluated vertices and labeled edges between the vertices. Each vertex of the semantic network represents a concept, object, idea, function, or activity, etc. The edges denote relations (links) between vertices. Such links can be used to derive new information and may form the basis for the inferential process. In semantic networks, the use of links in a reasoning mechanism is called inheritance hierarchy. This means that items lower in the net can inherit properties from items higher up in the net.

The problem in using semantic networks arises from the fact that the definitions of the concepts are subjective and difficult to incorporate into a program [13].

2.1.2 Frame representations

The term frame, in artificial intelligence, refers to a special way of representing knowledge and concepts. Both semantic nets and frames are considered as frame-based mechanisms. Similar to a semantic net, a frame representation is a structure of nodes and relations organized in a hierarchy, where the top most nodes represent general concepts and the lower ones represent more specific instances of the concepts [9]. However, in a frame system, the concept at each node is defined by a collection of attributes and values of these attributes, called slots. Each slot can have procedures attached to it which are executed when the information in the slot is changed In a frame structure, knowledge concerning individual objects is represented by instance frames. It has been suggested that descriptive knowledge can be successfully represented in a frame-based structure [14].

2.1.3 Production rules

The concept of using a rule format to represent heuristic, verbal knowledge originated in the DENDRAL project [15]. The rule format knowledge representation scheme uses conditional statements to represent knowledge and is probably the most popular type of knowledge representation technique In this format, the IF part contains the premise of the rule, and the THEN part indicates the action or conclusion of the rule. Well-written rules can be transparent and the developer of the rule base should be able to see through the syntax to the meaning

In a rule-based expert system, the domain knowledge is represented as a set of rules. Once the IF portion of a rule is satisfied by the facts, the action specified by the THEN portion is performed. Knowledge represented in a rule base can be hard to modify since rules are interconnected and changes of one or several rules may affect the entrie logic represented in the rule base. Another drawback of a rule-based system is that it is hard to use rule format to express structural knowledge

Commercially available expert system shells often come with fixed knowledge representation schemes that are more proficient only for certain types of applications. Theoretically, the possibility of using several of these knowledge representation schemes jointly may be appealing, however, the development of such an expert system shell, or in other words, the design of a standard knowledge representation scheme is difficult For practical use, the choice of a knowledge representation scheme should be as convenient and appropriate as possible [16].

2.2 INFERENCE MECHANISMS

The reasoning mechanisms of an inference engine are normally described as either forward chaining or backward chaining. Some inference engines also support a mixed chaining mode.

2.2.1 Forward chaining inference

Forward chaining is also called data-driven reasoning. In a forward chaining system, the inference engine starts by examining the rule's premise and fires those rules that are satisfied, this process runs successively until all the conclusions are identified based on the fact(s) provided by the user. The depth-first search mechanism implies that the search moves from one level of knowledge to another once a rule at that level has been fired Therefore, this mode of inference will be much faster than a breadth-first approach. However, it is possible that the inference engine may not find an answer, since the search strategy does not access all the knowledge at each level and may travel along a path on which there is no solution. The mixed chaining strategy can reduce this possibility.

2.2.2 Backward chaining inference

Backward chaining or goal-driven reasoning begins with a suspected goal and extends backwards from there in an attempt to find evidence that could match the hypothesis from data in the knowledge base. To avoid a dead-end solution, the system requires that the user supply the missing facts that are necessary for the inference engine to discriminate between responses so that a single goal can be reached.

2.2.3 Mixed chaining inference

Mixed chaining refers to an inference strategy that uses both forward and backward chaining mechanisms with a single knowledge base Normally, the mixed chaining inference process starts in the forward chaining mode and when the situation arises, e.g. insufficient evidence to satisfy a conclusion, the engine switches to the backward chaining mode that may allow a conclusion to be reached by asking goal-related questions.

2.2.4 EAshell: An expert system shell developed at U.W.O.

Expert systems developed based on forward chaining or backward chaining strategies can be combined with either a depth-first or a breadth-first search Depth-first searches require a search through all levels of a knowledge base until a conclusion or a contradiction is reached. In this process, the inference engine continues to pass from level to level based on a single affirmative answer. Breadth-first searches require that at each knowledge level all the premises must be searched thoroughly, followed by a new search of the next knowledge level. This process is repeated until a conclusion is reached. A depth-first approach can be faster, however, if the system gets stuck while going along a path on which there is no solution, this inference strategy is not able to resolve the dilemma since it only follows the path without cross checking the neighboring knowledge levels. Nevertheless, it is possible to combine a depth-first search with other search strategies to provide the possibility of choosing which path to explore. This is an effective modification which allows a knowledge-based program to display expertise at user-specfied levels without exhausting the system's resources.

EAshell is an expert system shell developed at U.W.O. that uses a rule-based knowledge representation method and supports all three of the above mentioned inference strategies within the Microsoft Windows© environment. In EAshell, the forward chaining mode takes a depth-first approach, while the backward chaining strategy utilizes a search approach rather similar to the breadth-first approach, and the mixed chaining mode makes use of a modified forward chaining process that provides the system with the flexibility to handle situations when more facts (missing facts) are necessary to complete an inferential process. Implementation of EAshell in the development of expert system applications will be presented in Chapters 4 and 5 of this thesis.

2.3 A DESCRIPTION OF KNOWLEDGE ACQUISITION

Typical problems tackled by expert systems are generally not procedural, instead they include information that is vague, complex, incomplete, or even incorrect [17]. Solutions to those problems demand a great deal of expertise. Consequently, knowledge-based expert systems need to use more than factual knowledge to solve problems that involve some degree of uncertainty. As outlined in Figure 2.1, the stages in the construction of an expert system include

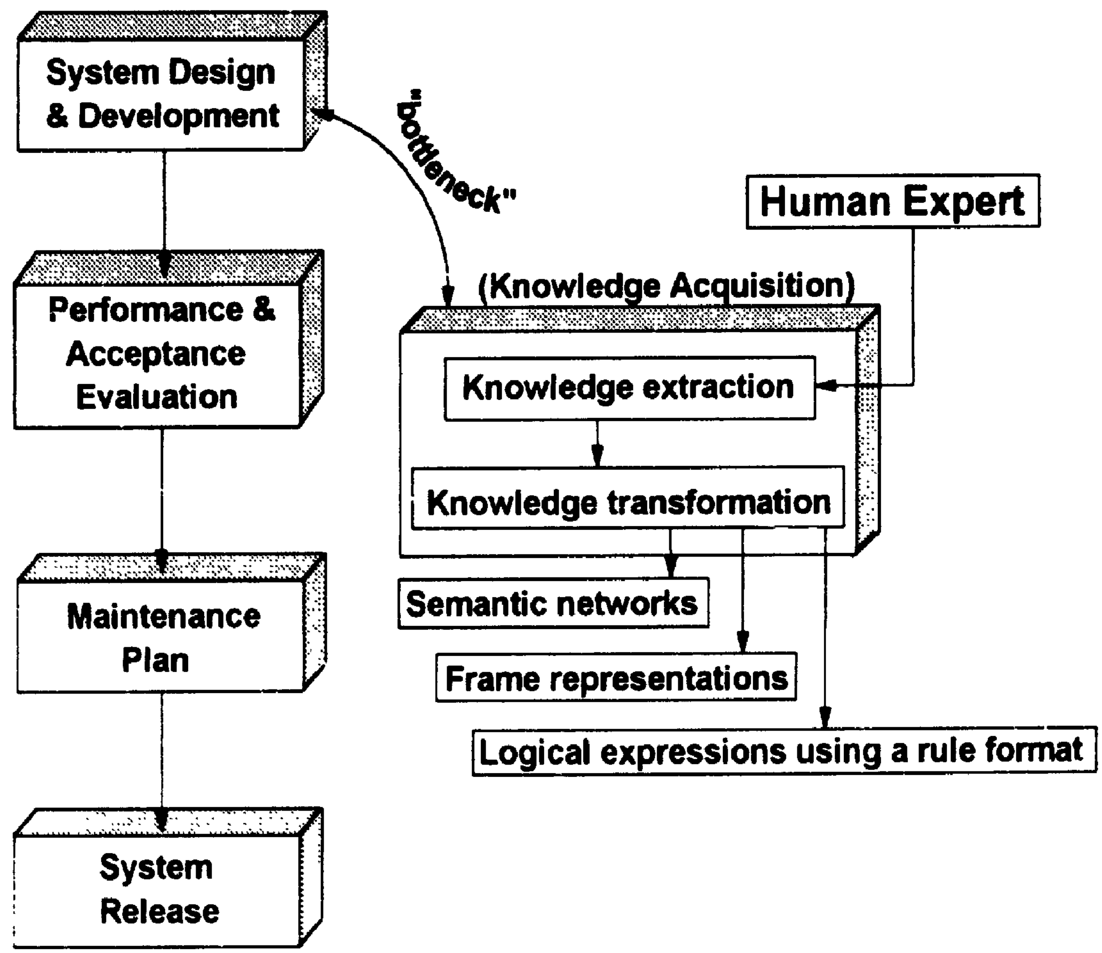


Figure 2.1 The stages involved in construction of an expert system.
Knowledge Acquisition is a particular important step that can be the bottleneck for the subsequent development work

system design and development, evaluation of performance and acceptance, maintenance, and finally, system release. Knowledge acquisition is an especially important aspect in the early stages of system design and the subsequent development Later, the knowledge base must be modified to correct ambiguities and be extended to reflect new knowledge.

2.3.1 Characterization of human knowledge

In almost every scientific specialty, human knowledge is divided into two categories, namely factual knowledge and heuristic experience [18]. Factual knowledge, which has been defined as support knowledge by Breuker *et al.* [19], is a category of knowledge that has been published in a well-written format, a system developer can acquire this type of knowledge through research. Private knowledge comprises heuristics (or rules of thumb) that human experts apply in manipulating and interpreting the factual knowledge. When faced with a practical problem, both the expert and novice soon find that solutions cannot be found solely using factual knowledge, special expertise or heuristic knowledge enters into the problem-solving stream.

Unlike factual knowledge, heuristic knowledge does not have a single well defined value and is hard to express and document precisely [20]. One way, and probably the only effective way, to become familiar with heuristic knowledge that is used in a specific domain of expertise is through application of the knowledge in problem solving processes. Experts are experts because they know both the factual and heuristic knowledge that defines their domains.

2.3.2 The bottleneck problem

Knowledge acquisition is the process that transforms expertise (both facts and heuristics) into implementation formalisms that can be understood by computer programs. This process is problematic and known as the 'bottleneck' problem mainly because of the following two reasons.

The first is the human factor. Initially, a domain expert is primarily proficient in a narrow domain, the expert's knowledge is often insufficient to meet all the requirements of an entire target domain. This problem can be overcome by using a team of experts, however, with an increased number of people involved, the job of coordinating, meeting, and documenting for the purpose of knowledge transfer rapidly becomes more difficult. A number of additional reasons that are associated with the human factor and complicate this task have been described by Edwards and Cooley: the lack of willingness to share expertise, the difficulties in expressing expertise verbally as it becomes more and more domain specific, and the fact that many domain areas are themselves fields of research that are still poorly understood [21].

The second reason is the software factor Verbal data do not speak for themselves, they have to be interpreted. However, there are no ready-made interpretation mechanisms available that can satisfy both the requirements of easy formalization and coherence. Although there are a number of expert system shells available that may reduce the amount of work required in the development of an expe system, these shells are themselves rather rigid systems that only support prefabricated knowledge representation mechanisms and may not match requirements of the target domain in which a knowledge-based system is under construction.

At present, many commercially available expert system shells adopt the decision-tree structure. Our own results and the reports from other research groups have suggested that the tree-like structure may be ineffective for knowledge encoding [20,22-23]. Not only are the logical relations presented in a tree structure rigid sc that any changes made to the tree may upset the total relationship, but also the connections in a decision-tree are obscure, so that a knowledge engineer may not be able to understand the structure easily The only feasible way to update a knowledge base developed in this way may be through a major revision from root to tip.

2.4 THE KNOWLEDGE DOMAIN MATRIX SCHEME

There are two important requirements for the form of knowledge representation employed in an expert system. First of all, the knowledge representation selected for an expert system should accountably represent what the human expert knows and how he uses that knowledge to solve domain problems. This does not mean that the chosen representation method has to be a psychological model, exactly imitating a human's reasoning process, but it does mean that the method must be able to capture the fullest possible range and power of the human expert's knowledge in that particular domain. Next, if an expert system is to be responsible for complex decision-making and giving advice, then it is vital that there is compatibility at the cognitive level, between the user's model of problem and the system's. In other words, the knowledge representation method employed by the system must be readily intelligible to the user. Only if this is true, will the user both be able to interact competently and efficiently with the system during its reasoning process and also be confident in the system's reasoning and advice.

In this section, we describe an alternative knowledge representation mechanism that has been developed and used in the development of a number of expert systems in our laboratory [23-26].

2.4.1 A causal knowledge model

If intelligent computer programs are to provide genuinely expert levels of performance in the problem-solving stream, they must incorporate some sort of causal model, both to support expert problem-solving and to provide an effective interface with the system developer as well as with the system user.

Thus, to understand how a human expert reasons about causal relations between his (her) observations and the subsequent decisions is of pragmatic importance in conceiving a knowledge model that may be useful to structure heuristic knowledge into a certain primitive form, and consequently facilitate the process of knowledge elucidation, formalization, representation, and ultimately utilization by computer programs.

Heuristic knowledge is developed by field experts through years of practice and is used in the manipulation and interpretation of factual knowledge to solve practical problems. The way in which an expert applies heuristic knowledge in pursuing answers towards a target problem may be represented logically by a causal analysis expression using conditions (OBServerables) and conclusions (ACTions). Such expressions can be represented as:

$$\text{Decision } \Psi(ACT_1, ACT_2, \ldots ACT_n) = \Phi(OBS_1, OBS_2, \ldots OBS_m) \ldots\ldots (2.1)$$

Where OBS_i ($i = 1 - m$) are observables and ACT_j ($j = 1 - n$) are actions of the decision-making process Ψ.

Sometimes, we observe that a set of conditions may result in a number of suggestions within which the inference engine is not able to discriminate. This incomplete set of conditions together with further observations may begin another round of decision-making process, and consequently produce a more precise decision. Such processes can be represented as:

$$\text{Decision } \Psi'(ACT_1', ACT_2', \ldots ACT_p') = \Phi\{\text{Decision } \Psi(ACT_1, ACT_2, \ldots ACT_n) \mid OBS_1', OBS_2', \ldots OBS_q'\} \ldots\ldots (2.2)$$

2.4.2 The Knowledge Domain Matrix (KDM) prototype

If heuristic knowledge can be interpreted by using the causal analysis knowledge model and expressed as a function of conditions and conclusions, it is possible to develop a tabular form to accommodate the results from causal analysis of a knowledge set. Such a two dimensional table is called a Knowledge Domain Matrix (KDM). In the KDM scheme, the causal analysis results are transferred into an empty matrix in which the conditions are listed as the first row in the matrix, and the conclusions in the first column of the same table. At the primary level, the logical connections between conditions and conclusions in a KDM are established by filling in "True" and "False" values in corresponding cells of the matrix according to the causal analysis results. At this stage, the logical connections are mainly arranged along the diagonal region of the matrix and only valid between individual cells.

The logical relations of the entire knowledge matrix do not exist until the next step, known as secondary knowledge encoding, is completed. At the secondary level, conditions and conclusions that were not related at the primary level are connected logically, thus the knowledge represented by a KDM is expanded to cover the off-diagonal regions of the matrix The knowledge in a KDM is said to be completed when all the conditions and conclusions are properly connected [23]

Unexpected conclusions may arise as a result of using a KDM based only on the primary and secondary knowledge compilation. This is due to the fact that sometimes "parallel cases" occur when one conclusion can give rise to two sets of conditions. In this situation, an additional layer of knowledge is added to the KDM at the tertiary level to differentiate the similarities between single conditions which contribute to the parallel cases.

Blank cells in a KDM represent a "no" connection, neither "True" nor "False", between specified conditions and conclusions. The domain expertise represented by a completely filled KDM is much greater than at the primary level because the knowledge contained in the off-diagonal regions (through both secondary and tertiary compilation) correlates the entire knowledge domain expressed in the KDM. Figure 2.2 shows a simple example of a KDM structure. More detailed discussion and examples of completed KDM can be found in subsequent chapters of this thesis as well as in publications by this group [20,23-27].

CONDITIONS / CONCLUSIONS	Ion trap contaminated, conditioning overnight	Electronic multiplier gain drop off	Ion trap temperature too low	Filament current too low
Low sensitivity		T	T	T
Noise decreases as column temperature decreases	F			
Short ionization time	T			F
Ionization time varies		T		

Figure 2.2 A KDM example that illustrates a small portion of the knowledge domain matrix used in the GCMSdiagnosis expert system.

2.4.3 The multi-layer knowledge domain matrix

In reality, however, the decision-making process may often require multiple inferential processes to be performed at different levels, and each of these levels may require the use of previous conclusions, together with additional observations as new conditions, to carry the reasoning process on further [20]. For instance, the diagnostic procedure necessary to solve a fault in an hyphenated instrument, such as GC-MS, requires at a minimum, a two-step inferential processes to be performed in tandem. In terms of the KDM, this multi-step inference process cannot be represented by a single two dimensional matrix. In Figure 2.3, we show a multi-layer matrix structure designed as part of this work to address such a complication in the knowledge domain. In this Figure, the domain knowledge is represented by separate knowledge matrices that each focus on a subdomain of the entire problem domain.

As a causal knowledge model the multi-layer KDM structure is designed to incorporate the knowledge complexity found in multi-step decision-making processes. In this scheme, conclusions from a previous knowledge level may be used as factual information in deeper levels of inferences. This structure is more useful to organize knowledge from complex knowledge domains in which sub knowledge layers exist. For example, in order to represent the knowledge related to the fault diagnosis of GC-MS analysis, the domain knowledge can be arranged into two knowledge layers in which the first KDM concentrates on diagnosis of analysis by gas chromatography, and the second focuses on diagnosis of MS-related problems. Results show that when a complex knowledge domain is organized in a multi-layer KDM model, it is more structured with narrow focuses set on sub knowledge domains, therefore, the representation of knowledge in a rule-based system can be more precise. In fact, determining an appropriate response to a chemical spill involves a multi-step decision process and we have developed a multi-layer KDM to represent the domain knowledge and used it as the knowledge base for the

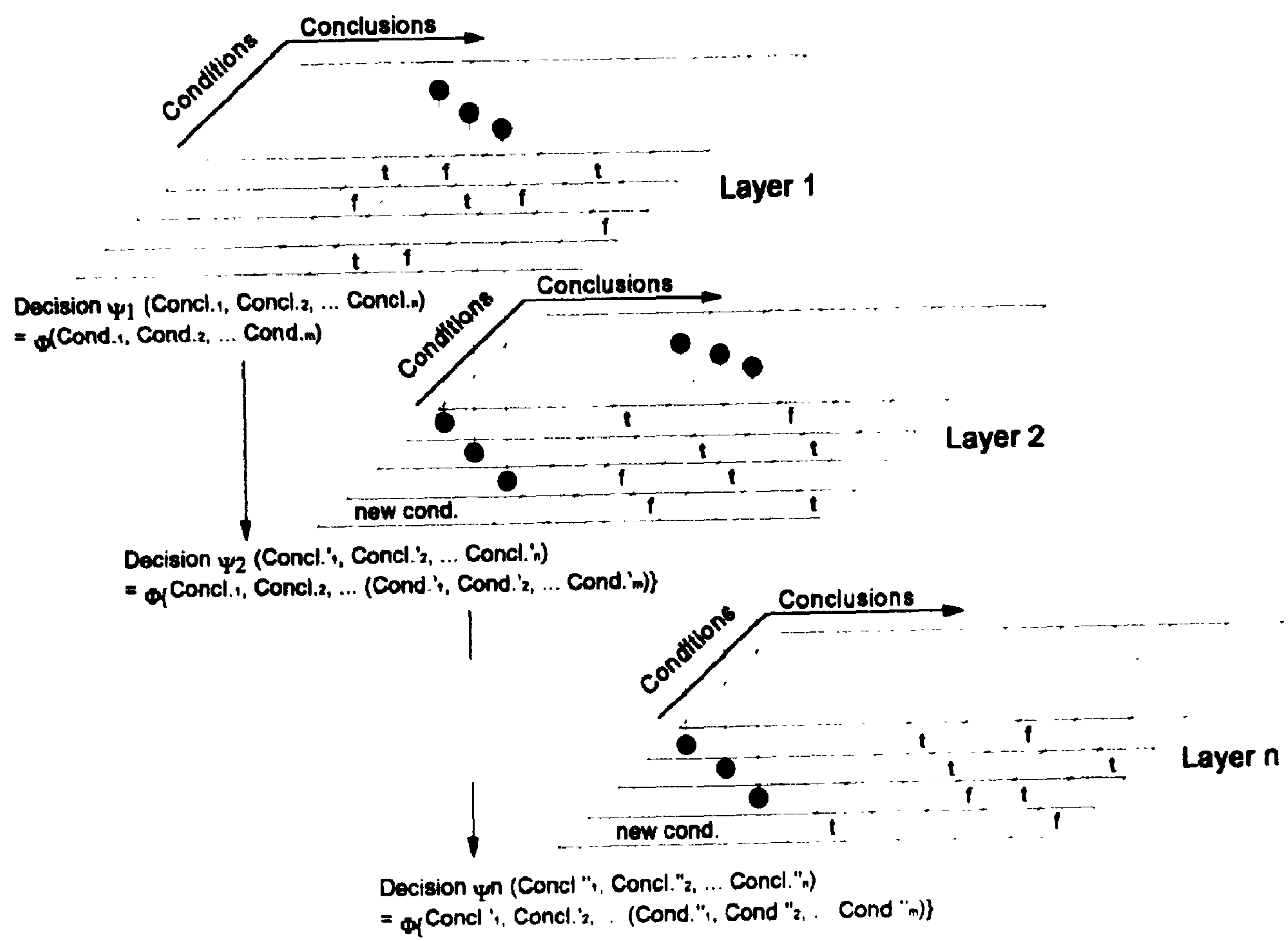


Figure 2.3 The multi-layer knowledge domain matrix structure.
The multi-layer KDM structure is designed to incorporate the knowledge complexity found in multi-step decision making processes In this scheme. conclusions from a previous knowledge level may be inserted as conditions in deeper levels of inferential processes

ERexpert program.

2.4.4 The rule-based knowledge file

The domain knowledge represented in a KDM needs to be transformed into a certain logical format before it can be used by an inference engine. In our projects, an inference engine developed at U.W.O. [28], named EAshell, is used as the core with which we build expert system applications. EAshell employs a rule-based knowledge representation mechanism, therefore, the knowledge acquisition results represented in a KDM need to be converted into production rules to enable the utilization by the inference engine.

In a rule format, the coded heuristic knowledge is represented by a series of IF condition(s) THEN conclusion(s) sentences. A tool kit, named RuleEditor that is part of the shell, has been developed to carry out the conversion of a filled KDM into a rule-based knowledge file (KBF) for use by EAengine automatically [28]. The resulting KBF file consists of three major sections: a goal section, a logical expression section, and a user query section. The goal section provides information about goals used for backward chaining and contains a set of goal-related variables. The logical expression part is the main rule-based section in which heuristics are represented in IF...[AND/OR]... THEN... rules, an explanation (*EXP*) statement is also given for the conclusions of each rule. Finally, the user query section provides questions to ask that are related to the facts and provides options as answers for identification. It is important to note that when both AND and OR are used in a rule, OR takes precedence over AND as if parentheses surround expressions connected by OR. The contents of a KBF file and a brief explanation are given in Table 2.1.

Table 2.1 shows the details of Rule 33 from the knowledge base used by the diagnostic expert system module developed for gas chromatography (GC) and tandem mass spectrometry (MSMS). Rule 33 fires when *high background noise*, *poor resolution*, and *variable retention time* are identified in a chromatogram. The conclusion suggests that the cause for such irregular chromatographic behavior may result from a high carrier gas flow. The explanation suggests a typical range of gas flows and how to adjust the gas flow.

Table 2.1 An Example of a KBF File and a Brief Explanation

KBF Example	Explanation
Goal Section **FIND** Leaking, Inject, Column, IonTrap, CarrierGas, Sample, Communication; :	Starts with ***FIND*** and ends with ";". Each Goal Variable represents a sub area (part) of the instrument, rules are organized around the goal variables.
Logic Expression Section RULE 33 IF Variable1=YES **AND** Variable17=YES **AND** Variable18=YES; **THEN** CarrierGas=Carrier gas flow too high; **EXP** "Optimal gas flow rate is between 0.5 to 2.0 ml/min. Adjust the control on the GC instrument and make sure the gas flow stays within the range." :	This block begins with keyword ***IF***, the premises are connected by ***AND***s or ***OR***s, and the pointer ";" ends the block. ***THEN*** indicates the conclusion of a rule. ***EXP*** gives a full explanation of the result immediately after that rule block.
User Query Section **ASK** Variable1 "High background noise masks sensitivity?" **OPTION** YES, NO; **ASK** Variable17 "Unresolved peaks?" **OPTION** YES, NO; **ASK** Variable18 "Retention time change from run to run?" **OPTION** YES, NO; :	The ***ASK*** keyword begins a query block while a ";" ends it. Corresponding content of each goal variable is given by the ASK sentence in the Query Section. ***OPTION*** clause provides available answers to questions.

2.4.5 Maintenance of the knowledge base

Maintenance of an expert system, namely the need to update the knowledge base to include more knowledge and to delete redundant knowledge, is an important part of the life of an expert system. It is especially important in the early development stages that the knowledge base undergoes frequent modifications and expansions to ensure correct answers are provided. Results from using commercially available expert system shells seem not to be very promising under the perspective of both our own observations and the reports from other research groups [22,23]. As described earlier the causes for this problem are that the "decision-tree" structure adopted by many expert system shells makes maintenance difficult. Moreover, the connections in a tree-structure are behind the scene, so that the knowledge engineer is not able to see the connections, which in turn makes a change in the logical relations more difficult. The KDM process, which was adopted for this work, uses a two dimensional matrix to hold the results of the causal analysis for a given knowledge domain. Instead of using a tree-like structure, the logical relations in a KDM are made by filling in "T(rue)" and "F(alse)" values to corresponding cells in the matrix. Unlike a tree-structure in which the logical relations between conditions and conclusions are "hard-wired", the logical connections in a KDM can be edited flexibly by assigning new logical values to those cells. This type of modification is independent of the existing relationships so the structure of the KDM will not be affected. The unique advantage of the KDM scheme is that it enables the knowledge engineer to see directly the logical relations held in the knowledge matrix, so that a knowledge base developed in this way can be re-structured and expanded quite easily [23,27].

Hypermedia is an alternative way of implementing and structuring knowledge in an expert system shell, and has recently received much attention in the development of knowledge-based systems [29]. We find the hypermedia process interesting because it provides looser connections between the conditions and conclusions so that the system user can freely change, add or delete nodes of a decision tree developed using such methodology. However, it is problematic because a naive user will find it difficult, if not impossible, to use such a knowledge system in an effective manner. In our opinion, such freedom should only be given to the knowledge engineer during the system development and upgrading processes.

2.5 CONCLUSIONS

The slowest step in the development of a knowledge-based system is the step involving knowledge elucidation and representation. In this chapter, we have explored the difficulties associated with the knowledge acquisition process, and developed a causal knowledge model to be followed in analysis of domain knowledge. Based on this causal knowledge model, a novel knowledge encoding method is evolved. Expansion and update of the knowledge base of an existing expert system is an important step in expert system development cycle, however, the constraints involved in adding new knowledge into the knowledge base make such modification very difficult. This difficulty can be largely eliminated with the use of the KDM scheme. As discussed in this chapter, a knowledge base developed through the use KDM can be easily modified.

The results discussed in this chapter suggests that the KDM process is a promising knowledge encoding method that can match the knowledge encoding requirements from a complicated domain problem, although further verification will be required to test this methodology in more applications

2.6 REFERENCES

1. Lederberg, L.; Sutherland, G. L.; Buchanan, B. G.; Feigenbaum, E. A.; Robertson, A. V.; Duffield, A M.; Djerassi, C. *J. Am. Chem. Soc.*, **1969**, 91, 2973-2976.
2. Duffield, A. M.; Robertson, A. V.; Djerassi, C.; Buchanan, B. G.; Sutherland, G. L.; Feigenbaum, E. A.; Lederberg, L. *J. Am. Chem. Soc.*, **1969**, 91, 2977-2981.
3. Esteban, M.; Ruisanchez, I.; Larrechi, M. S.; Rius, F. X. *Anal. Chim. Acta*, **1992**, 268, 95-105.
4. Buydens, L.; Schoenmakers, P.; Maris, F.; Hindrinks, H. *Anal. Chim. Acta*, **1993**, 272, 41-51.
5. Maris, F.; Hindriks, R.; Vink, J.; Peeters, A.; Driessche, N. V.; Massart, L. *J. Chromatogr.*, **1990**, 506, 211-221.
6. Schoenmakers, P. J.; Dunand, N. *J. Chromatogr.*, **1989**, 485, 219-236.
7. Luinge, H. L. *Vibr. Spectrosc.*, **1990**, 1, 3-18.
8. Koutny, L. B.; Yeung, E. S. *Anal. Chem.*, **1993**, 65, 148-152.
9. Waterman, D. A. *A Guide to Expert Systems.* Addison-Wesley Publishing Company: Reading, Massschusetts, 1986.
10. Kidd, A. L. In *Knowledge Acquisition For Expert Systems: A Practical Handbook*; Kidd, A. L., Ed.; Plenum Publishing Co.: New York, 1987; Chapter 1, p 1.
11. Ringland, G. A.; Duce, D. A. *Approaches to Knowledge Representation: An Introduction.* John Willey & Sons Inc.: New York, 1988.
12. Kidd, A. L. *Knowledge Acquisition For Expert Systems. A Practical Handbook.* Plenum Press: New York, 1987.
13. Mac Randal, D. In *Approaches to Knowledge Representation-an introduction.* Ringland, G. A.; Duce, D. A., Eds.; Research Studies Press Ltd.: Letchworth, Hertfordshire, England, 1988; Chapter 3, p 45.

14. Lucas, P.; van der Gaag, L. *Principles in Expert Systems (2nd Edition)* Addison-Wesley Publishing Co., Workingham, England, 1991, Ch4, pp173-252.

15. Feiganbaum, E. A.; Buchanan, B G. *Artif. Intell.*, **1993**, 59, 233-240.

16. Buydens, L.; Peeters, A.; Massart, D. L. *Chemometr Intell. Lab Syst*, **1988**, 5, 73-79.

17. Pierce, T. H.; Hohne, B. A., Eds.; *Artificial Intelligence Application in Chemistry.* American Chemical Society: Washington, DC, 1986; ACS Symposium Series 306.

18. Hayes-Roth, F.; Waterman, D. A.; Lenat, D. B , Eds.; *Building Expert Systems.* Addison-Wesley: London, 1983: Vol. 1.

19. Breuker, J.; Wielinga, B. *in Knowledge Acquisition For Expert Systems. A Practical Handbook;* Kidd, A. L., Ed.; Plenum Press: New York, 1987 Ch 2, p17.

20. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 945-955.

21. Edwards, M.; Cooley, R. E. *CABIOS*, **1993**, 9, 657-665

22. Penninckx, W.; Smeyers-Verbeke, J.; Massart, D. L.; Spanjers, L. G. C. W.; Maris, F. A. *Chemometr. Intell. Lab. Syst.*, **1992**, 17, 193-200.

23. Lahiri S.; Stillman, M. J. *Anal. Chem.*, **1992**, 64, 283A-291A.

24. Du, H.; Lahiri, S.; Huang, G.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 21-31.

25. Zhu, Q.; Du, H.; Plomely, J.; March, R.; Stillman, M. J. GCMSdiagnosis - An expert system for diagnosis of problem GC-MS data. *Anal. Chim. Acat*, **1996**, submitted.

26. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 956-968.

27. Du, H.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 33-41

28 Huang, G.; Stillman, M. J. An expert system shell under development in our laboratory, unpublished results.

29 Bourguignon, B.; Vankeerberghen, P.; Massart, D. L. *J. Chromatogr*, **1992**, 592, 51-57

CHAPTER 3
THE ROLE OF THE USER INTERFACE IN EXPERT SYSTEM APPLICATIONS

3.0 INTRODUCTION

Within the academic world, recent achievements in the field of expert systems are remarkably impressive. At the same time, it is disappointing how few successful working systems exist outside research laboratory environments To improve this poor track record, a major effort must be committed to improve the user interfaces of those knowledge-based systems. As Kidd and Cooper pointed out, the effectiveness and acceptability of an expert system is critically dependent on its man-machine interface [1].

The evolution of expert systems is making interface design a more critical factor in the overall development of an application. It is obvious that in order to make man-made machines to come up to a level where they can 'replace' human experts, it is also necessary to equip these machines with facilities that could allow close interactions between the machines and their users

There is no point in having a powerful knowledge-based system if users cannot easily and efficiently communicate with it under the constraints of the real-time task. It is a real challenge facing present-day expert system developers to develop capable user interfaces that can match the dialogue facilities of such applications with the communication needs and constraints imposed by the class of users for the systems and the types of tasks in question For example, most expert systems require a very flexible mixed-initiative dialogue, so that the system users can easily take over control from the system at any point and volunteer information or request an explanation, etc

The emergence of high-performance personal computers with graphical display functions is having a considerable impact on work in expert system development. The availability and capability to develop software in multitasking environments, such as Microsoft© Windows, in which both data and objects can be displayed graphically and commands can be issued by pointing and clicking rather than by typing, has certainly reduced the difficulties in development of graphical user interfaces to meet the above-stated requirements.

In this chapter, I emphasize the important role of the user interface in the development of expert system applications, some basic concepts about the user interface are to be addressed together with discussions of the results achieved in design of a Microsoft© Windows based graphical user interface that is generally implemented in the applications developed in our laboratory.

3.1 DEFINE USERS OF EXPERT SYSTEMS

The user interface requirements for expert systems have evolved considerably since the days when a consultation system first conducted a dialogue with its user. Such changes are due in part to the transition from expert systems to expert advisory systems, the increasing use of deep models of knowledge, and the increasing size and complexity of the systems [2]. Early expert systems operated in a way in which they asked users many questions and then returned an answer. Differently, an expert advisory system user is actively involved in the decision-making process. In an expert advisory system, a system user needs to make decisions based on information and helps provided by the advisory system, that is, the user and the system share the reasoning and decision-making tasks. Because the user guides and participates in the reasoning process there is increased necessity for the user interface to support the user's cognitive task. Both the quantity and quality of communication between the system and the user need to be increased.

Generally, there are two sets of users for an expert system, and the design of the user interface should reflect the differences in their needs. The first type of users are consumers of the knowledge-based systems, such as a less experienced analytical chemist, who tries to use an expert system as a guide to perform required analytical tasks that demand domain knowledge beyond his/her knowledge realm. The second set of users are domain experts who use the system to examine and augment the knowledge base. Both kinds of users need the capability to interact with expert systems in problem solving processes, however, the expert users may also need to have an overview of the entire knowledge base so that the knowledge modification and expansion can be approached more systematically. In the development of the user interface for an expert system, both types of the system users need to be taken into consideration.

3.2 THE USER INTERFACE

Until recently, researchers in the field of expert system have focused their attention on identifying and solving topics related to how a computer system can be designed to solve specific problems in carefully selected, and well-defined problem domains. However, of the three primary elements of an expert system shown in Figure 1.2 illustrated in Chapter One, the user interface remains the least researched and developed. We believe that this component remains a hurdle to full acceptance and wide spread of knowledge-based systems used in business and industry.

To properly state the role of a user interface in an expert system shell, the key components of an expert system are rearranged into Figure 3.1, in which the importance of a user interface is clearly expressed. The user interface is the component of an expert system that allows bi-directional communication between the program and its users. At the most primitive level, the user must be able to pose a problem to the system and stimulate the system to respond with a suggestion. Furthermore, the system should be able to describe or explain its reasoning to the user who is uncertain of why a particular question(s) has been

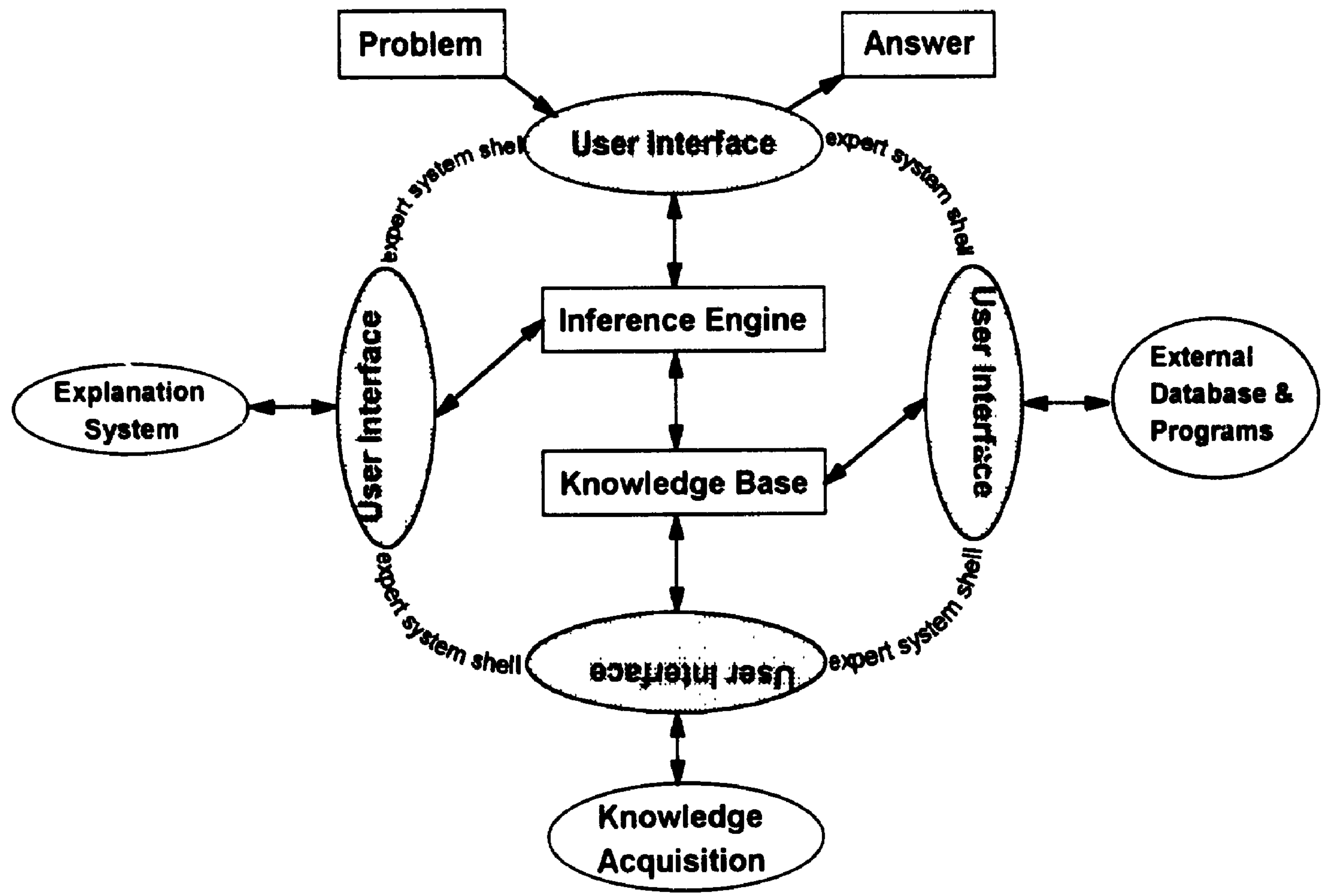


Figure 3.1 Components in an expert system shell.
There are three key components in every expert system: (1) the user interface, (2) the inference engine, and (3) the knowledge base. The user interface is the connection to each part of an expert system. In graphically-based expert systems, the user interface will incorporate pictures, images, and sound clips.

posed to him or her. Additionally, the interface maybe required to manage and display other types of information, such as graphics, video images, or sound clips. Knowledge acquisition for expert systems is a difficult and time-consuming process, very little is known about how to extract expertise from an expert and almost nothing is on offer as a technique to aid in the process. The problem is characterized on the one side by a domain expert, unfamiliar with expert systems and unable to articulate what knowledge he has and how he uses it to solve problems; and on the other side by a knowledge engineer who may well be ignorant about the domain of expertise. Therefore, a complete user interface needs also to provide certain interactive programs that can communicate with expert(s) to provide some kind of help during the knowledge acquisition process. Presently, the common input mechanisms or devices that enable this communication include the keyboard, mouse, and touch screen.

3.2.1 Requirements for the user interface

Requirements for interfaces to expert systems should be categorized into those supporting system developers and those supporting end users, although the distinction may often overlap. This is because an expert system shell needs to be modified and customized to meet the application demands and the knowledge base of a system requires updating to reflect newly arrived knowledge of the field, or to delete redundant information, and in many cases, the system users are actively involved in this process.

Theoretically however, the difference in interface requirements stem from the different focuses of the two classes of users. The major focus of a system developer is on the effectiveness of domain knowledge representation by the machine (computer), and the correctness of reasoning using such encoded knowledge in various inferential processes; on the other hand, an end user is focused on the problem-solving capability of an expert system and the convenience of using such a system .

For designing the user interface to an expert system application, there are six basic requirements that need to be considered (Figure 3.1).

1. The user interface should represent the domain in the user's natural idiom, this implies a need for user interfaces to provide interactive displays that can reflect the mental model of system users in their problem solving processes.
2. The user interface should provide immediate feedback on the system state to the users by explicitly maintaining and displaying complex constraints and interrelationships. This requirement demands the use of graphical functions to aid the envisionment process.
3. To recover easily from trying different alternatives implies more than just undo facilities, it represents the unique capability of expert systems to 'guess' answers based on the encoded knowledge. The user interface to an expert system therefore, shall be designed to reflect and augment such recoverability
4. The user interface shall allow system users to work at appropriate difficult granularities. Dedicated expert system applications need to have interface designs to decompose the complexity of target problems and help system users to understand each components and their relationships in order to facilitate problem-solving process.
5. The user interface must be implemented in such a way that supports multiple accesses to various information resources. Expert systems need to use both factual and heuristic knowledge to solve domain problems, and sometimes, external databases and functional procedures are also required.
6. Optionally, the user interface may be furnished with aid(s) to the process or knowledge acquisition and modification. Such facilities provide users with interactive programs that enable a systematic, formatted approach to knowledge acquisition.

These considerations of the user interface need to be judged against application realities. While a design of the user interface to an expert system shell may need to consider all of the above mentioned requirements, interfaces to small and dedicated expert systems can support only part of those previously described functions.

3.2.2 Types of user interface

New software shells and machine capabilities keep expanding the array of choices with which the designer is faced. In designing a user interface, many issues need to be considered, including problem type, modality, conceptual models, tasks, and problem-solving approaches [3]. Generally, there are four types of user interfaces or approaches to interface design.

Command-based interface Command-based interfaces were the 'original' type of user interface. To communicate with a computer through this user interface, a user needs to type in a code-like command or string of commands in required syntax according to the machine's resident language. This type of human-computer interface requires more of the human than the computer. Consequently, a system user has to be an expert in the command language in order to maximize the power of a computer program since no aids are presented to the system user during the communication.

Natural language interface Natural language interfaces are those that allow users to communicate with a computer in common language. These systems are also known as natural language front ends that include both natural language understanding and generation capabilities to enable users to give a command and read system feedback easily. A natural language program can work well if the terminology, phraseology, usage, and common request in the domain are well-defined and known to user.

However, natural language programs can not interpret every word entered correctly since human language is often ambiguous. For example, we use in our conversation words like 'average', 'lots', 'few', 'many', and 'several', these terms are not acceptable by machines since knowledge-based systems do not deal well with impression. To what numbers or range of numbers do these terms relate? Even humans may disagree among themselves. Moreover, humans use additional forms of communication to get their message across, including tone and non-verbal gesture, which a machine can not understand.

Menu-based interface Menu-based interfaces provide the users with a selective set of options from which to choose, a descriptive title that defines why the choices are being made, and an easy selection mechanism. Menus are a viable interface style for expert system because they reduce the need for the system user to memorize complex commands, reduce training cost, and enable systems to be used more efficiently by average users.

Well designed menus do restrict the total set of choices, but provide clear leads to navigate through a system. Users, especially those novice and infrequent system users can greatly benefit from the structure imposed by menus since they have less to consider and fewer possible mistakes to make. On the other hand, menus may be less effective with expert users as they may become agitated when required to make selection after selection to reach their goal state. It is frustrating to know what to select and not be able to find it. For this consideration, a makeup technique, called menu bypass, is developed. Macro functions are built into the systems that enable expert users to quickly access particular screens or goal state that otherwise may require a number of actions before they can be reached.

Graphical user interface Graphical user interface (GUI) is a subset of the direct manipulation interface. For a graphical user interface, the following guidelines are applicable: (1) the domain is represented by objects and actions,

(2) point, drag, and click replaces key presses as a means of communication, and (3) feedback is fast and actions are rapid and reversible. GUIs are currently the most popular interface style and this approach has been adopted widely by different computers and operating systems, such as Macintosh, MS Windows©. In the next section, a MS Windows© based GUI developed by our laboratory as a general user interface to our expert system applications will be presented.

There are a number of advantages of using a GUI as user interface to expert systems. First of all, pictures can be worth a thousand words. In many situations, the GUI can be the most effective choice that reduces cognitive demands related to operating the system and frees users to concentrate on problem-solving processes. Secondly, the graphical design can result in a more friendly environment in which users may feel less intimidated and more willing to explore functionality of a system. And finally, graphics enable the use of supporting communication elements such as animation and sound.

The graphical user interface may be less effective when there are extensive options available during a problem-solving process. This is because the task of searching through and translating graphical images become too cognitively perplexing, which in turn may interfere with the task at hand. Also, the GUI will not be effective if a domain is one in which there are so many possibilities that the designer may never able to ascribe icons for precise representation.

In practice, an expert system is designed for users of varying levels of expertise, and the nature of the problem domain differs considerably from case to case, therefore, the user interface to an expert system shall be designed accordingly. In many situations, different types of user interface styles are necessary to represent domain tasks in a way that encourage effective problem solutions. In fact, the combination of a GUI with command structures in different types of menus has been successfully demonstrated by Macintosh Windows technology.

3.2.3 Benefits of a good user interface

Expert system may be well suited for many programming situations, but an interactive system must have a good user interface to succeed [4]. The simple question and answer dialogue style can be inappropriate in many applications where greater visibility, user control, and user initiative are required. A good user interface can mean the difference between a 'working prototype' of a complex system and one that actually works as an effective advisory tool in a domain [5].

Benefits of a good user interface are summarized as following:

1. A good user interface can increase user acceptance of an expert system. Users with knowledge of a task domain may find the system easy to use.
2. A good user interface may increase the frequency of use of an expert system. Users with little computer concepts need only learn a small number of operational rules to be able to use the program comfortably. So that they can concentrate on their tasks and the computer seems to vanish.
3. A good user interface also means lower operational error rates, increased performance speed, and decreased training time for its operator. Not only are syntax and computer-domain errors reduced, but task-domain errors can also be eliminated in many cases. Users of such a system may feel freer to explore novel "what-if" possibilities without fear the system may hang-up on their actions. Even if contradiction does happen, a good user interface can provide an easy way out and restore to the previous level without losing much of the information.

3.3 A WINDOWS BASED GRAPHICAL USER INTERFACE DEVELOPED AT U.W.O.

The breadth and complexity found in chemistry means that a successful expert system application developed in this domain needs to be capable of interacting with several different modules, some are rule-based, some are graphically-based, and some are procedurally-based. The expert systems developed in our laboratory perform an advisory role during an analytical or environmental chemical problem-solving process that requires flexible yet effective communication between users and the systems. For example, during such a problem-solving process, the system may require information about the task, results may need to be displayed graphically, users may ask the system for explanation, and sometimes external information sources are necessary. Therefore, as part of the project in development of expert system for chemical applications, we need to devise a user interface that can be capable to handle the high degree of communication requirements as well as to maximize the information flow.

During the user interface design stage, we discovered it is difficult to achieve the required flexibility, simplicity, and effectiveness of this user interface. The complexity of a rule-based expert system tends to limit novice users from effective use of the program, on the other hand, expert users may be more satisfied when questions and answers are offered in a style that is equivalent to their knowledge of the domain.

To satisfy such communication requirements, we knew from the beginning that an object-oriented graphical user interface in combination with a menu command structure would be the best solution to our interface design. The difficulty then was the choice of a suitable programming environment that supports graphical and menu-item computing with the possibility of access by physical actions, such as point and click, drag of mouse, as well as key-strokes, yet the programming itself should not be too difficult to be used effectively in the development. Our major focus was to devise expert systems that can be used to solve chemical problems rather than to learn computing language(s).

We experimented with several feasible options that were then available to us The use of a commercial expert system shell imposes restrictions on modification of its existing front-end [6-8]. On the other hand, programming in the DOS environment is the most common practice and is relatively familiar to us, however, DOS applications, such as Quick Basic 4.0 (Microsoft Corporation), have limited graphical abilities and is hard to develop programs that could fully respond to mouse actions. Following a number of attempts, we decided to program in the Microsoft© Windows 3.x environment (Microsoft Corporation). We first used Actor 3.0 (Whitewater Group), an object-oriented programming language to write source code for the user interface, and soon discovered that the level of computing was too high and the documentation of the program was too poor to provide efficient programming guidance. Then we chose to use Visual Basic (Microsoft Corporation), a computer language that also supports the object-oriented programming structure and operates in the MS Windows environment. The unique advantage of using MS Visual Basic is that many commands that are used by the DOS version of the Basic language are also supported by this Windows version. To a large extent, the learning curve for using this language efficiently for the interface development is therefore reduced significantly.

Microsoft© Windows 3.x provides an operating environment that allows applications to be operated simultaneously with shared hardware resources, such as memory, key[illegible] mouse, screen, and printer, etc. When programming in MS Visual Basic, software developers do not need to write specific code to enable mouse actions or pull-down menus since such communication functions are embedded in Windows applications. Features provided by the Windows' Application Programming Interface (API) include:

- Shared display, memory, keyboard, mouse and system timer;
- Device independent graphics; and
- Multi-tasking abilities supported by functions such as, dynamic data exchange (DDE) object linking and embedding (OLE), dynamic linking libraries (DLL), and multiple-document interface (MDI)

We have developed a Windows-based graphical user interface for the ACexpert project (the ACexpert user interface). Through the implementation of Windows' characteristics such as described above, the ACexpert user interface provides effective communications for users to interact with application programs. In the following, the important features of this graphical user interface are outlined:

1. The ACexpert user interface encompasses multiple windows that can be accessed simultaneously. Figure 3.2 illustrates several images of this graphical user interface in which the multiple windows of ACexpert user interface are shown. Users can easily access different modules by point-and-click sequences through program and function icons, menus, and graphical images.
2. The ACexpert user interface consists of integrated control facilities, users of the program are able to use pull-down menus, icons, graphics, dialogue boxes, mouse, and keyboard to effectively facilitate the communication. As shown in Figure 3.3, the interface allows a combination use of point-and-click with key-press in a random order to select options to run an application.
3. The ACexpert user interface provides a highly intuitive graphical environment, users of the program have the flexibility to choose different media to view data or display inferential results. For example in Figure 3.3, the same data file is displayed numerically as well as graphically
4. The help facility of the ACexpert user interface, as demonstrated in Figure 3.4, provides users with on-line help once clicked, which can

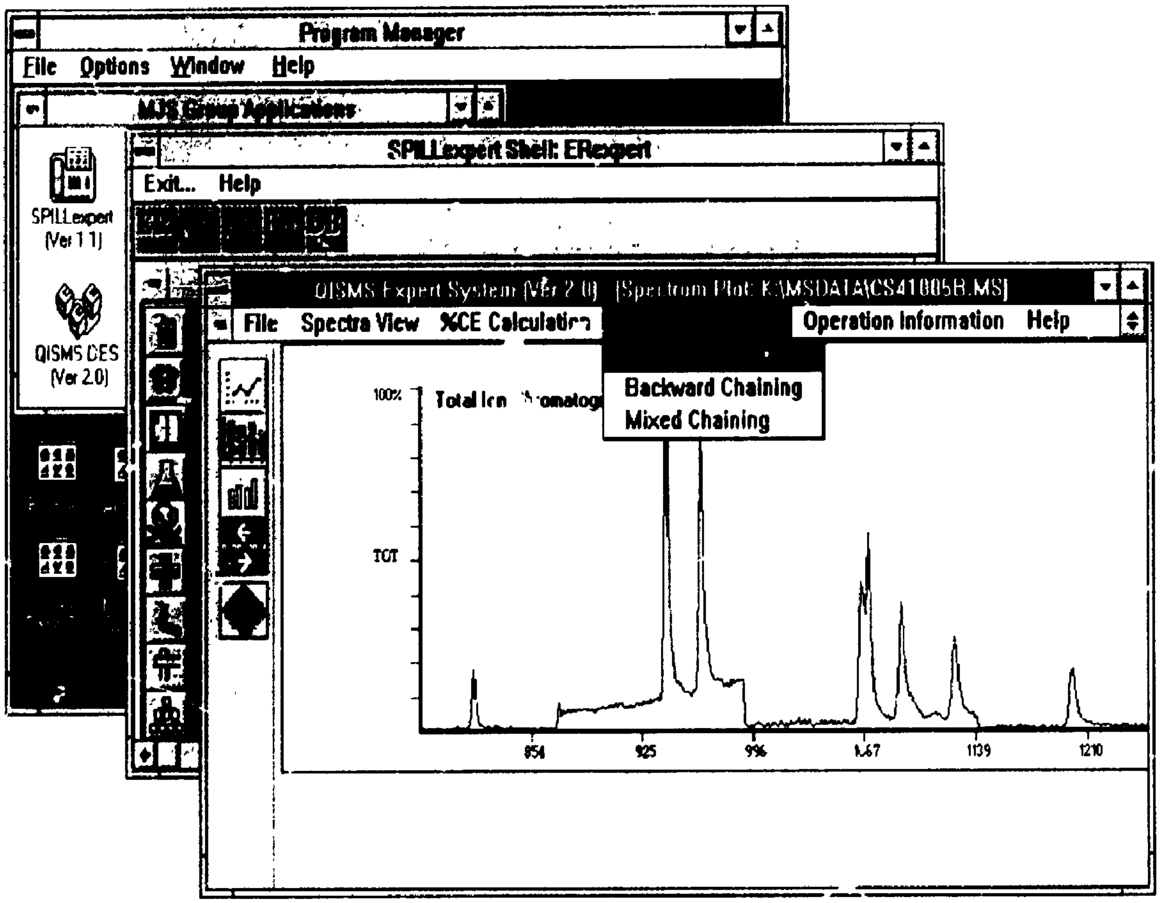


Figure 3.2 The ACexpert graphical user interface.
Multiple windows of the ACexpert user interface are displayed. Users can easily access different modules by using point-and-click sequences. Program and function icons, menus, and graphical images of the user interface provide users with easy accessibility to operate the program.

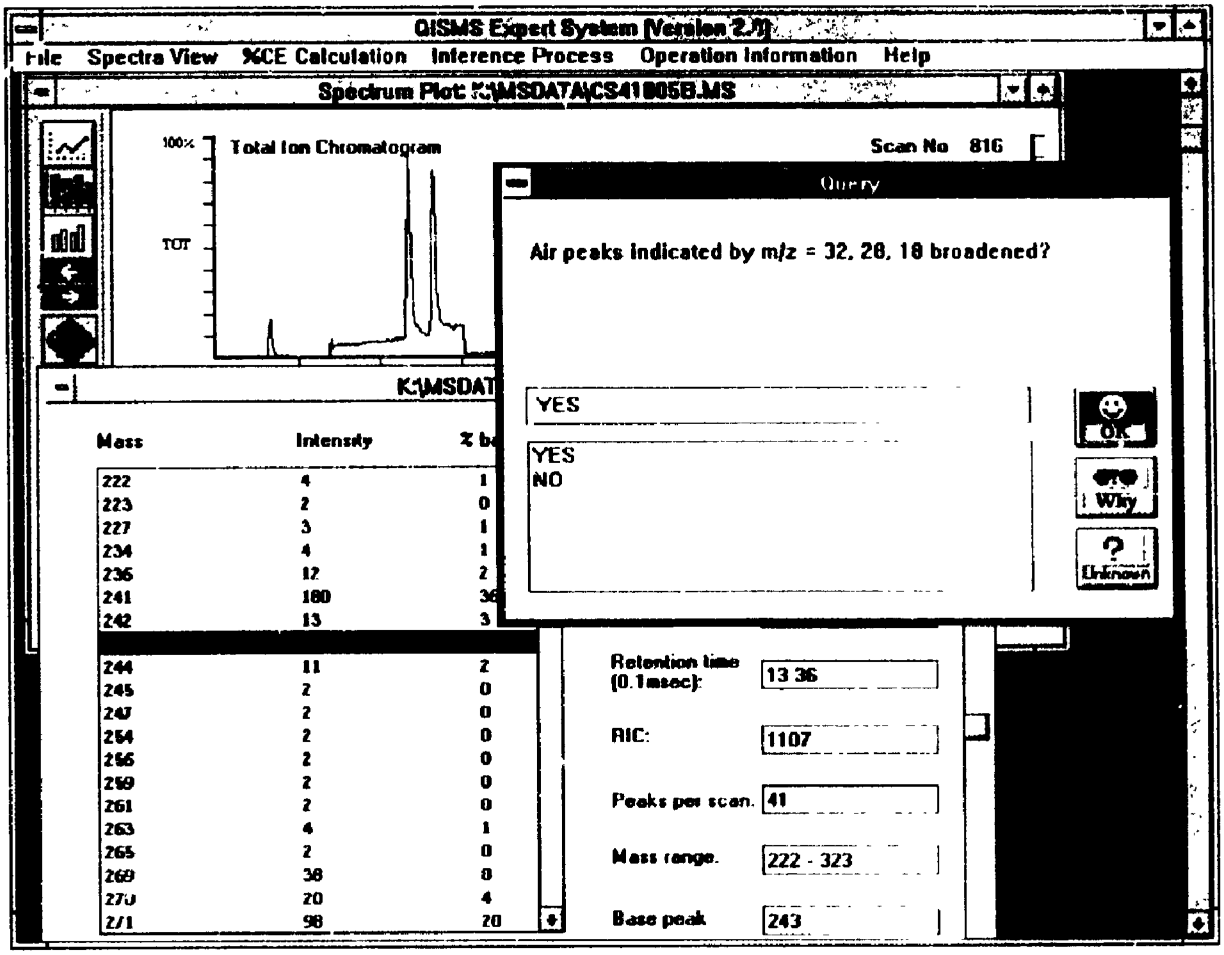


Figure 3.3 Integrated control facilities provided by the ACexpert user interface.

Users of the program are able to use pull-down menus, icons, graphics, dialogue boxes, mouse, and keyboard to communicate effectively with the program

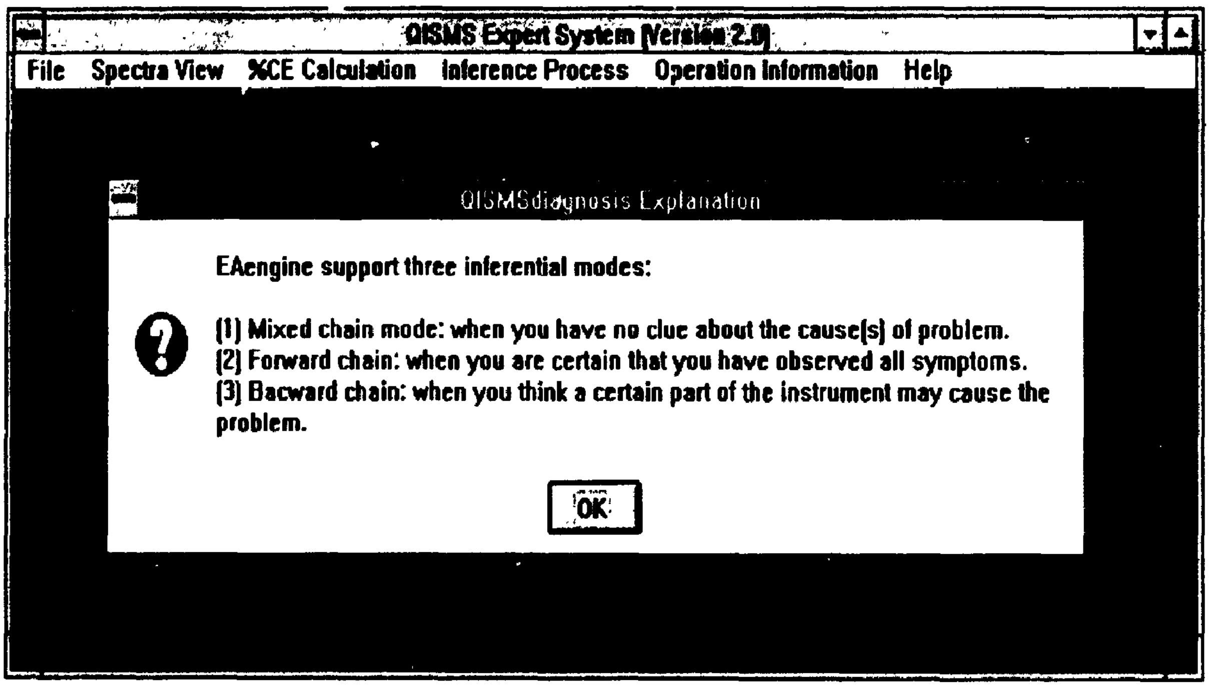


Figure 3.4 The help facility of the ACexpert user interface.
The 'help' function provides users with on-line helps once clickeu, which can facilitate the use of the program by delivering instant explanations for functions, questions, and results of the system.

facilitate the use of the program by delivering instant explanations for functions, questions, and results of the system.

The above illustrated general design and functionality of ACexpert user interface has been implemented in the interface design for a number of expert system development projects [9], including AAexpert [6,10], GCexpert [8,11], SPILLexpert [7,12], and GCMSexpert [13] systems. We have successfully implemented the characteristics of the graphical user interface in our own application programs. Presently, we are in the transition period, two versions of the ACexpert user interface are currently under implementation, one has been developed using MS Visual Basic, the other using MS Visual C The C code user interface is a modification based on the Basic code module, and because of the intrinsic power of the C language, the C code user interface is more efficient to execute, and certain limitations imposed by the Basic language can be overcome. However, we found that the use of MS Visual Basic in development of the graphical user interface is easier and adequate to cover requirements of general applications. Because the code writing in Basic is straightforward, the system developer(s), usually not a computer expert, can learn the programming quicker.

3.4 CONCLUSIONS

The knowledge-based program is characterized by its ability to deal with incomplete and imprecise information and to accumulate knowledge, it is essential to equip such an expert system with a powerful user interface so that system users can focus on solving domain tasks. Applications of expert systems within the chemical domain fall into two broad categories: (1) systems that are developed to interpret experimental results; and (2) systems that provide guidance as to the course of actions the analyst should take Both of them need to be furnished with graphical user interfaces in order to be practically operable under the time constraints of real application situations.

The user interface of an expert system is not the paint put on at the end of a project, but the steel frame on which to hang the details of a program. "Know thy user" should always be the guideline for user interface design. Exploring the methodology for an effective user interface design and the subsequent development of a sufficient user interface for the ACexpert project has been part of the research covered in this thesis. The choice of proper computing language is critical in the planning stage [14]. It took us quite sometime to try out different programming languages, such as Quick Basic (Microsoft Corporation), FORTRAN (Microsoft Corporation), and ACTOR (Whitewater Group), as well as commercial expert system shells, like KDS (KDS Corporation) and Cexpert (Software Plus, Ltd.). However, these systems were either too rigid, with very limited graphical capabilities and programming flexibility, to meet the basic requirements of our graphical user interface design, or the level of programming required was too high so that programs could not be easily written by chemists.

The development of the user interface for an expert system must start with some concept of the structure of the domain expertise, how human experts perform the domain tasks, and how the results should be presented. At this stage, attention must be directed to the primary objective of analysis of the domain knowledge structure and a blueprint for subsequent development of the expert system application must be made. Drifting away from this primary goal, such as spending too much time on code writing, will cost the overall quality of the program.

Using MS Visual Basic, we have successfully developed a graphical user interface for the ACexpert project In the following chapters, further results from implementation of this generic user interface in the development of modular expert systems within the ACexpert project will be presented.

3.5 REFERENCES

1. Kidd, A. L., Cooper, M B. In *Principles of Expert Systems*, Gupta, A , Prasad, B. E., Eds., IEEE Press. New York, 1988; p319
2. Stelzner, M.; Williams, M D. In *Expert Systems: T' ꜝ User Interface*, Hendler, J. A., Ed.; Ablex Publishing Corporation. New Jersey, 1988, Chapter 12, p. 285.
3. McGraw, K. L. *Designing and Evaluating User Interfaces for Knov.'edge-Based Systems*. Ellis Horwood: New York, 1992
4. Baroff, J.; Simon, R.; Gilman, F.; Shneiderman, B In *Expert Systems. The User Interface*; Hendler, J. A., Ed.; Ablex Publishing Corporation New Jersey, 1988; Chapter 6, p. 99.
5. Hochman, Z.; Pearson, C. J.; Litchfield, R W Agriculture Syst , **1994**, 44, 217-235.
6. Lahiri, S.; Stillman, M. J. *Anal. Chem* , **1992**, 64, 283A-291A
7. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 956-968
8. Du, H.; Lahiri, S.; Huang, G.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 21-31.
9. Zhu, Q.; Stillman, M. J. Expert Systems and Analytical Chemistry Recent Progress in the ACexpert Project. *J. Chem. Inf. Comput Sci.*, **1996**, in press.
10. Lahiri, S.; Yuan, B.; Stillman, M. J. *Anal. Chem.*, **1994**, 66, 2954-2963
11. Du, H.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 33-41
12. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 945-955
13. Zhu, Q.; Du, H.; Plomely, J.; March, R.; Stillman, M J. GCMSdiagnosis - An expert system for diagnosis of problem GC-MS data, *Anal Chim Acta*, **1996**, submitted

14. Zhu, Q.; Stillman, M. J. ACexpert: Design of a Graphical User Interface. In *Proceedings of the 74th Canadian Chemical Conference and Exhibition*, Hamilton, Ontario, June 2-6, 1991; 61 AN-E1.

CHAPTER 4
SPILLexpert PROJECT: DOMAIN DEFINITION, KNOWLEDGE ACQUISITION, AND MODULE DEVELOPMENT

4.0 INTRODUCTION

For any chemical spill accident, the potential for environmental damage and time are linked through an exponential relation that puts a premium on response practices that can be enacted immediately [1]. This requires both highly efficient reporting systems and quick deployment of a complex set of countermeasures; yet each spill is an individual situation. In practice, not only are these requirements often beyond the capability of a single human expert, but also because of the unpredictable nature of an accident as to time and location, experienced personnel and the necessary referential information may not be available. On the other hand, computer-based expert systems can provide availability and consistency together with increased task performance and are, therefore, ideal tools with which to address complicated problems that require prior knowledge to be applied under stress and time constraints, such as in planning the response to a chemical spill.

Kollig *et al.* [2] pointed out that the quickest and least expensive way of obtaining environmental information is by accessing a database. Such databases may be capable of furnishing the user with literature reports, however, in many cases the problem is that the lack of consideration of the environmental reality makes such information hard to be implemented easily in a real situation. Literature data and previous case-history records cannot be applied under new circumstances unless the heuristic knowledge of the field experts on how to use these data can be captured and re-applied properly to interpret the difficulties found in the re-application processes

The specific actions required for the clean up of chemical spills are determined by a combination of the physical, chemical, geological, and biological properties of the released compounds in the complex matrix of the natural environment. It is imperative that information used for planning such a response is valid for the micro environmental surrounding the spill. However, a literature search revealed that little human expertise has been documented that can be used to help in solving the difficulties regarding the use of the rather massive amount of information that is available on techniques, regulations, and procedures involved in the response to a chemical spill. Certainly, the lack of effective software that can be used to capture such complex heuristic knowledge is one of the reasons why the extrapolation of available spill response technology to the environmental reality is extremely difficult.

By integrating the knowledge of the domain expert about a rather narrow field of study into a computer program, expert systems can be designed to provide specific advice based on incomplete and uncertain information [3-6]. In the environmental area, expert systems did not appear until the mid 1980s. The slow emergence of environmentally related expert systems is primarily due to the lack of a well-established scientific foundation for environmental science. However, because few environmental problems can be solved using expertise from a single knowledge domain, knowledge acquisition and subsequent coding also greatly inhibited development. Thus the difficulties in trying to incorporate the domain knowledge into an expert system to deal with environmental problems are much more significant [7]. In 1987, Hushon identified 21 environmental expert systems [8]. This number has doubled in the last two years [9]. Examples of these systems include the US EPA's environmental monitoring system, which was developed to increase the accuracy, timeliness, and cost-effectiveness of field sampling, chemical analysis, and data validation [10], the GEOTEX expert system for hazardous waste site investigations [11], the ES-EPA program for environmental pollutant analysis developed in Japan

[12], and a risk assessment model for chemical constituents of hazardous waste [13]. More recently, the Soil Treatment Evaluation Program (STEP) was developed through the use of object-oriented programming techniques [3]. STEP provides a utility that facilitates the selection of preliminary screening technologies applicable in the treatment of hazardous-waste-contaminated soils. Hanratty and Joseph have reported a comprehensive research effort to capture the knowledge used in the selection of laboratory reactors [14]. The computer-aided response technologies selector (CARTS) is an expert system designed to assist in designing the treatment train, identifying data requirements, and allowing users to evaluate different scenarios [15]. The remedial action assessment system, a computer methodology developed by Buelt *et al.*, aims at estimating remedial alternatives in terms of effectiveness, applicability, and cost [16]. However, to our knowledge applications in the field of emergency response to chemical spills have not yet been reported.

Defining the most effective methodology for the development of expert system applications in a multidisciplinary domain is challenging. In this chapter, we first focus on the description of the domain problem, then on the classification of the decision making hierarchy and tabulation of the domain knowledge, and finally on the implementation of the knowledge acquisition results. The progress made in the design and the subsequent implementation of a novel knowledge acquisition scheme in the development of a modular expert system has been fully documented. **ERexpert** is the emergency response expert system designed to offer advice following a spill accident. The prototype has been constructed using a minimal set of toxic chemicals to provide proof of concept. The development of **ERexpert** involves the following components: (i) the implementation of the KDM mechanism to allow a more efficient transfer of heuristic knowledge to production rules for use by the expert system [17-18], and (ii) the development of a database structure that provides an effective archival method for organization of the elaborate factual information in the domain. The

ERexpert program is a knowledge-based program that uses both static and conceptual data to support the strategies to be followed by the user in order to select the most favorable response countermeasures using supplementary information required for the chemicals involved in the spill. A graphical user interface has been developed that provides essential communication between the system user and the knowledge base. Finally, the possibility of integration of the **ERexpert** module into the **SPILLexpert** frame is discussed.

4.1 DOMAIN PROBLEM DESCRIPTION

4.1.1 Chemical spills

Hazardous materials, i.e., chemicals that are toxic, corrosive, flammable, or explosive, are a ubiquitous aspect of modern life [19]. In today's industrialized world, tens of thousands of different chemicals are produced each year. Many of these substances are not unusual materials utilized by only a few special industries, rather they are raw materials used by many manufactures in their every day production activities. These materials are manufactured and stored often in ton quantities by the producers; transported by truck, airplane, train, ship, barge, or pipeline; stored by the purchasers; and used in many manufacturing processes. At any point along this chain, a mishap may occur resulting in a hazardous materials incident. While prevention should remain as the main defense against environmental disasters, spills will inevitably occur. Accordingly, contingency plans have been and must continue to be developed for use in response to such accidents. Computer technology, especially knowledge-based expert systems, will play an increasingly important role in the environmental problem-solving process.

4.1.2 The current situation in Canada

In Canada, an average of two accidents occur every day involving the transportation of dangerous goods. Since 1985, Federal Transportation

Dangerous Goods legislation has required companies to have an approved Emergency Response Assistance Plans (ERAPs) in place. These plans provide documentation which enables industry to supplement the normai capabilities of emergency response teams. Some industries have gone beyond the requirements of the law with voluntary initiatives. The Canadian Chemical Producer's Association (CCPA) initiated a proactive program called "Responsible Care". Participation in the program is a condition of membership in the association in which the producers are taking responsibility for their productions and downstream use of the products by customers. One dimension of the program is the requirement that member companies develop emergency response plans to deal with accidents involving their products. In most cases concerning large companies, in-house response capabilities are in place. Outside private emergency response contractors with special training and equipment augment these in-house capabilities. On the other hand, smaller firms lack the financial resources to have in-house teams and rely more heavily on outside contractors. In either case, awareness of the need for emergency response capability and the use of private contractors has spread beyond the chemical manufacturing community to other sectors that handle or use potential dangerous materials.

A recent report [20] cited the statistics released by Transport Canada indicate a decreased trend nationwide in the total number of accidents occurred in 1989 and 1993 (see Table 4.1). Unfortunately, there has been no system to coordinate emergency response between various private contractors, even though the overall response system relies on them to a greater and greater extent. Nor is there an objective to assist client industries in evaluating the real capabilities of such contractors beyond their stated claims

Canada needs a more coherent national capability to response to chemical spills so that resources can be used more efficiently, expertise can be more accessible, and responses can be more precise and consistent. Clearly, the current system needs improvement in order to evolve into an emergency response network in which computer technology, especially those knowledge-based expert systems will play an indispensable role.

Table 4.1 Total number of chemical spill accidents by mode of transportation or by provinces in Canada in 1989 and 1993.

Accidents: Mode of Transportation (1989 & 1993)					**Accidents by Province (1989 & 1993)**				
	1989	%	1993	%		1989	%	1993	%
					Alberta	176	19	140	21
Road	252	27	228	35	BC	132	14	90	14
					PEI	3	0	3	0
Rail	64	7	35	5	Manitoba	41	4	56	8
					NB	62	7	15	2
Air	6	1	8	1	Nova Scotia	52	6	21	3
					Ontario	272	29	173	26
Marine	8	1	0	0	Quebec	120	13	108	16
					Sask.	36	4	38	6
Facilities	603	65	389	59	Nfld.	28	3	6	1
					N.W.T.	7	1	5	1
					Yukon	2	0	5	1
Total	933	--	660	--	Total	933	--	660	--

Source: Transpot Canada

4.1.3 Definition of the response paradigm

Decisions are commonly made based on a set of criteria established according to a priority list. This forces the expert to focus on smaller but more tractable sub-problem domains, because these subdomains are less complicated and better defined. Figure 4.1 describes the general aspects involved in the response to a chemical spill incident. There are three major factors to be considered in evaluating a possible protocol to be used in response to a chemical spill: (i) the scientific and technological responses, (ii) the regulatory constraints, and (iii) the political, economic, and social demands.

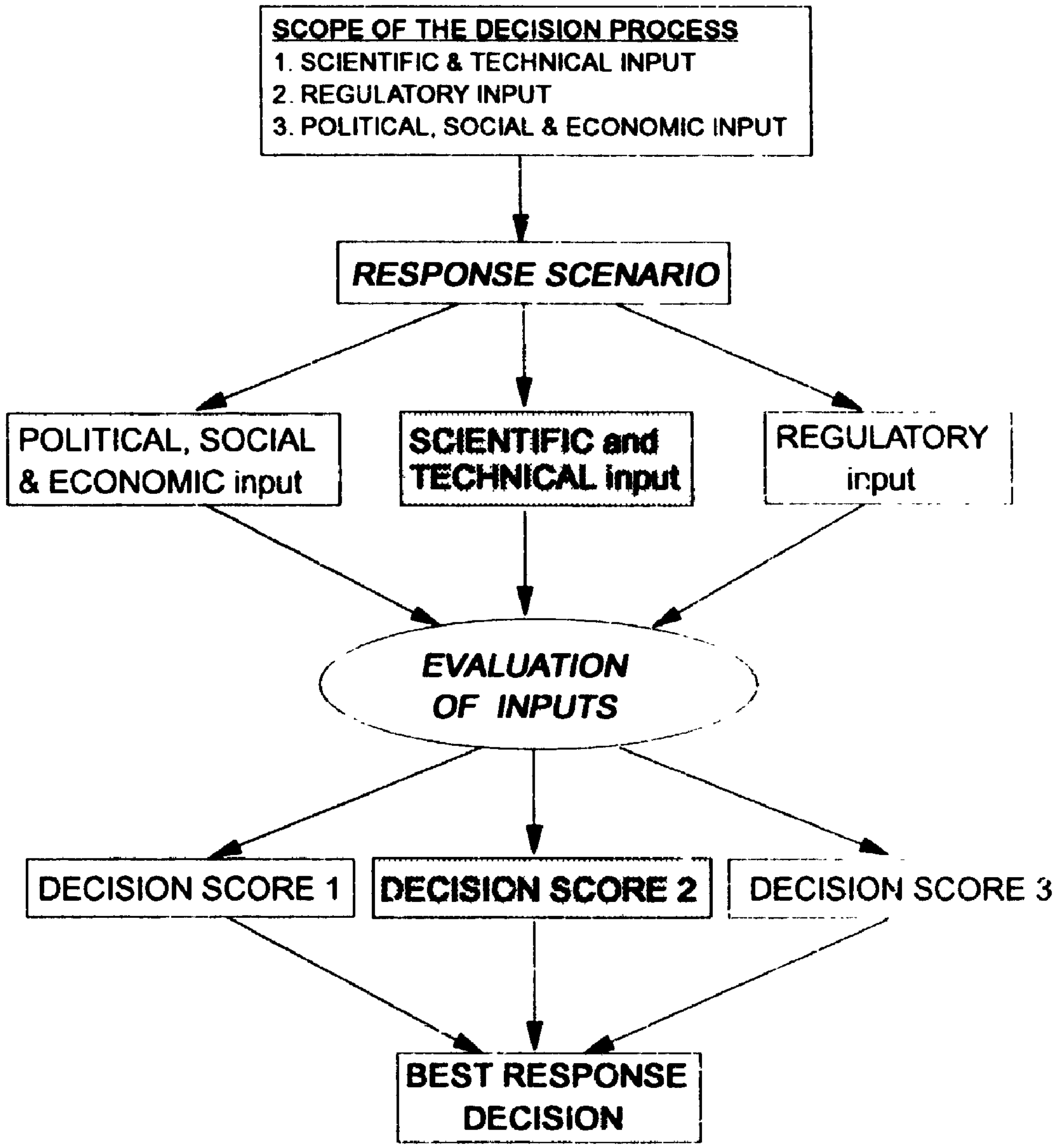


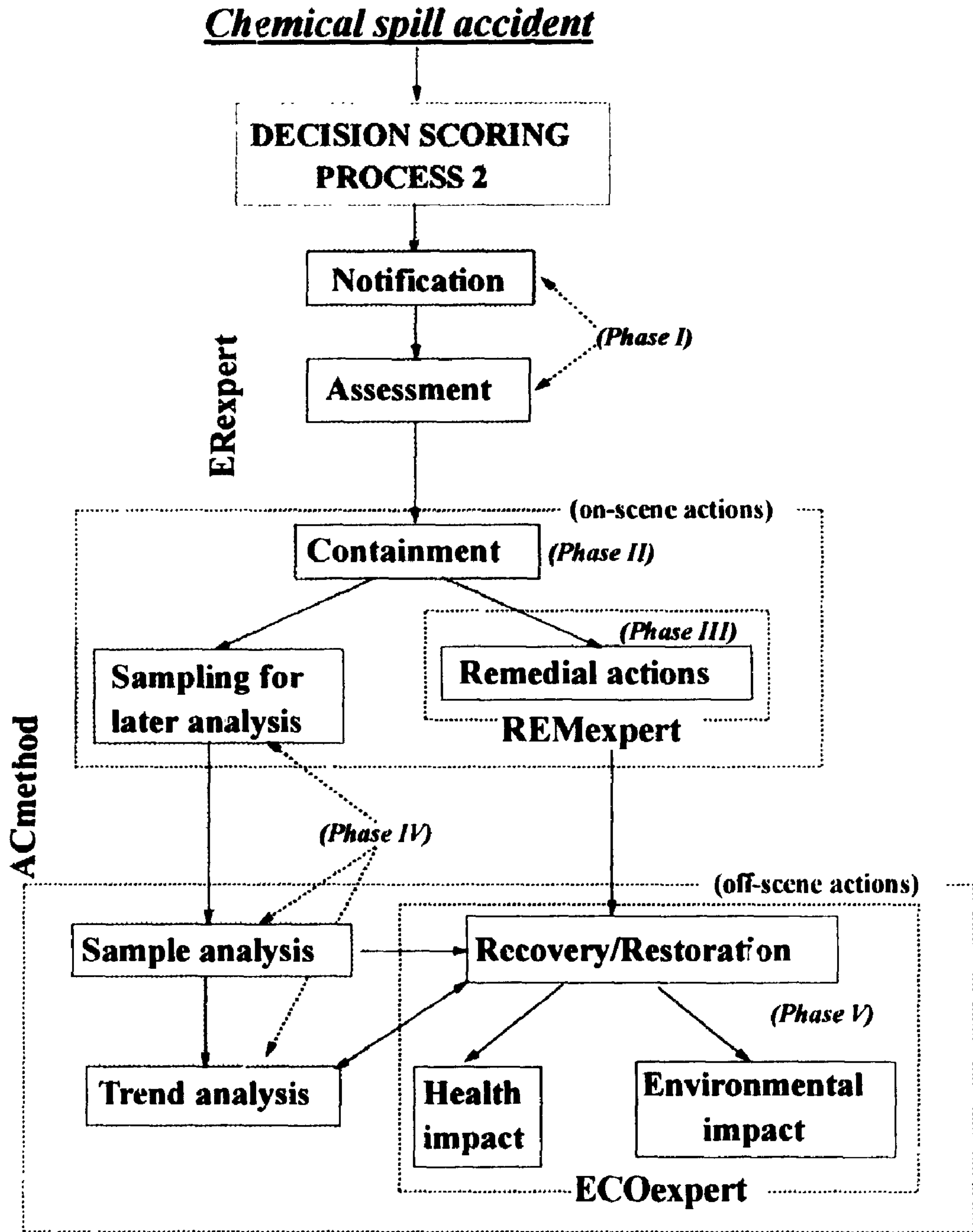
Figure 4.1 A description of the different aspects involved in response to a chemical spill accident.
Three major factors are to be considered: (i) scientific and technological response, (ii) regulatory constraints, and (iii) political, economical, and social demands.

This description is consistent with the comments made by Breshears *et al* [21], and Levin [22]. It is obvious that a reasonable response protocol can only be constructed by considering a combination of all of these factors. In order to reduce the complexity involved, the development of the prototype expert system described in this chapter has focused on the knowledge subdomains of the scientific and the technological aspects that covers response to a spill. However, despite these imposed limitations the knowledge base of the system must still incorporate expertise from a number of scientific disciplines

4.1.4 The conceptual decision-making hierarchy

Establishing a conceptual decision-making hierarchy (the order of actions) that represents the problem-solving process for the target is an important step before knowledge acquisition can begin. The decision-making hierarchy outlines the logic flow and depicts possible aspects that must be considered in solving a particular domain problem. This process is described as the "knowledge orientation" step [23]. The purpose of this step is not to acquire the actual knowledge but rather to elicit the knowledge structure necessary to develop a conceptual model that the knowledge engineer can follow in the subsequent design of an application. Figure 4.2 shows a decision hierarchy that can be followed in response to chemical spills. This decision hierarchy consists of five relatively independent components arranged according to the order of actions taken.

1. Notification and Assessment is the first step taken in response to a chemical spill. A rapid and correct assessment is essential because the extent and the potential damage of a spill provides information that is critical in the following phases of reactions. In some cases the response in this step might be decisive for the success of subsequent actions
2. Containment is the next and may be the most critical step in the response to a chemical spill. The aim of this step is to limit the extent of the spillage to the smallest possible area through deployment of properly



Note: sub problem domains in which modular expert systems are under development in this laboratory.

Figure 4.2 The decision-making hierarchy applicable in response to a chemical spill accident.

From the scientific and technological point of view, five phases of actions can be outlined in a spill response process. Each individual phase of actions is relatively independent, yet as a whole, they complete the technical requirement in response to a chemical spill.

chosen containment techniques. In both these phases, time is of particular importance, any delays in assessing a situation or implementing correct countermeasures may turn an accident into a disaster.

3. Remedial action, as the third step, includes physical removal and chemical treatment of the contaminants from the environment and the safe disposal or treatment of all collected hazardous materials. While cleanup of a contaminated environment is the real essence of the remedial action, in an emergency response to a chemical spill, it is of secondary importance because only after achieving success in the first two phases, can the remedial actions effectively minimize the environmental damage.

4. The sampling and Trend analysis phase is used to evaluate the success of the deployment of the previous actions. The data acquired in this phase can be used to direct subsequent actions and to document the incident for reference in the event of similar cases.

5. Recovery and Restoration is the final phase of the response and during it actions are undertaken to restore the environment to its pre-spill conditions.

It is important to keep the hierarchy as simple as possible while maintaining a desired level of coverage for the problem domain. Shank and coworkers suggest that a small set of primitive actions will account for most of what must be represented in the physical world [24]. Similarly, Zhou *et al.* grouped the activities in a standard analytical laboratory into families of primitive actions and subsequently developed a robotic system to perform these operations [25].

4.2 COMPUTER ENVIRONMENT

All the computer work was carried out on an Intel 80386, 20 MHz computer, with 8 megabyte of RAM, two 150-megabyte hard disks, a 1.44-megabyte floppy disk, and a super VGA graphic monitor.

The **ERexpert** program is designed for use in the Microsoft Windows© 3.1 environment. Software chosen for the development of this project are all MS Windows© based applications. Microsoft ACCESS© (ver 1.0), an interactive relational database system was utilized for factual data management. EAshell© is a Windows based expert system shell developed in our laboratory that provides the inference engine [26]. Microsoft Visual Basic© (ver 3.0) was used as the programming language for the development of the graphical user interface; the inference processes are initiated by calls to EAengine which is compiled as a dynamic linked library (dll).

4.3 THE STRUCTURE OF THE SPILLEXPERT PROGRAM

The stages involved in construction of an expert system have been defined as: system design and development, performance evaluation and acceptance, and system maintenance and release [27]. The knowledge acquisition step is key in the design and development stage. We have found that the knowledge acquisition process is most effective when it is focused on particular cases rather than on the entire target domain. We have successfully implemented this concept in the development several expert system modules to address consultation and controlling tasks in analytical procedures [28-38]. Such cases involve relatively narrow, more specific and, therefore, less complex problem subdomains. In the **SPILLexpert** project, we experimented with this approach and applied it to a much broader and complex target domain. The subdomains involved in response to a chemical spill have been described earlier (Figure 4.2). Consequently the global expert system structure that is designed must encompass three frames. Each of the frames in Figure 4.3 comprises several members that are themselves stand-alone expert systems performing independent tasks and are connected with the main goals of the program.

As depicted in Figure 4.3, **SPILLexpert** is the substructure corresponding to the scientific and technological subdomain that encompasses (i) **ERexpert**,

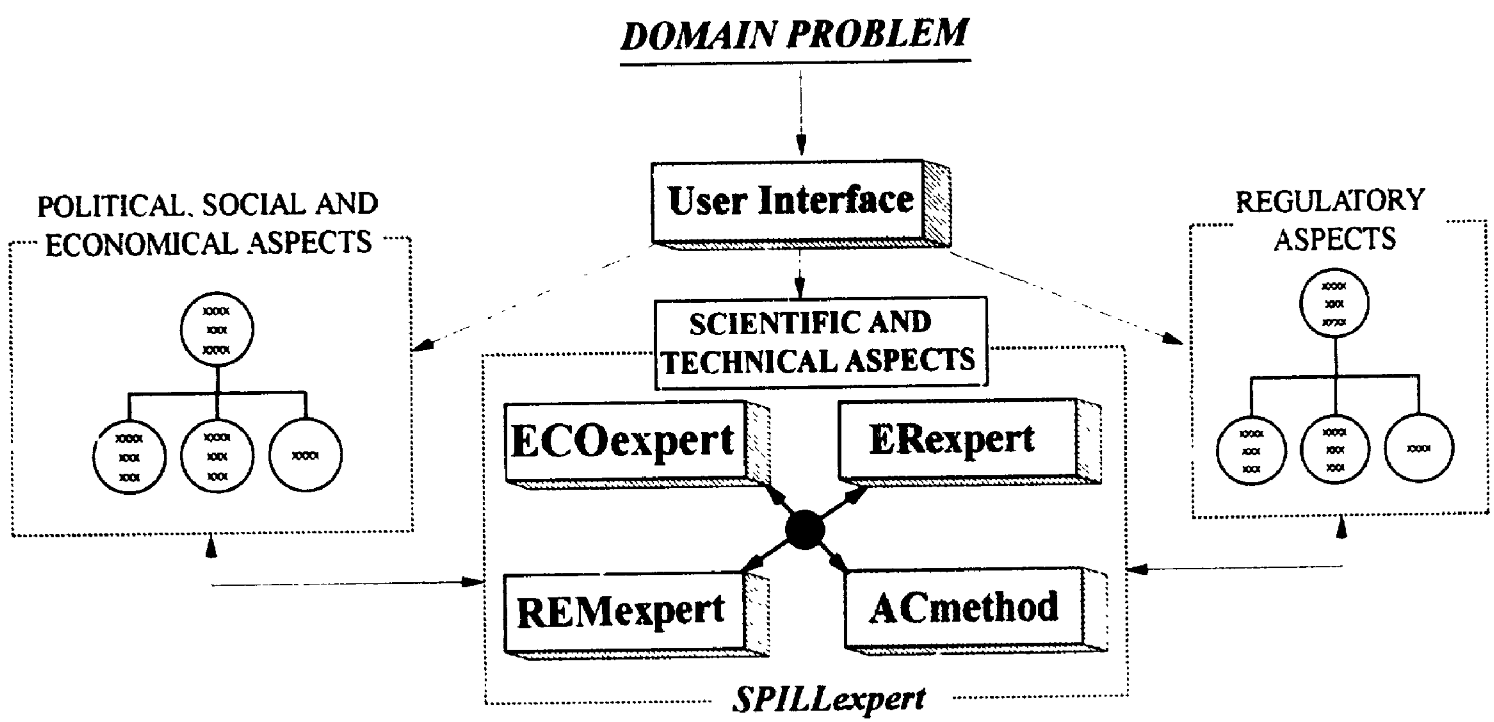


Figure 4.3 General structure of the SPILLexpert program.
SPILLexpert is a three-frame network in which a number of subsystems are attached to each of the frames. Development has been focused on the ERexpert and ACmethod modular expert systems, and the user interface (shaded areas) that are part of the project and described in this thesis.

(ii) **ACmethod**, (iii) **REMexpert**, and (iv) **ECOexpert**. **ERexpert** is a modular expert system developed to plan emergency response following chemical spills. The design is emphasized on the containment of the spill and deployment of remedial techniques during the primary phase of the cleanup. **ACmethod** is used to select appropriate sampling and analytical methods based on the matrix, sample types, and sampling conditions. In conjunction with **ERexpert**, **REMexpert** provides remedial action instructions for the next phase of the cleanup once the situation is under control and the incident has been contained. **ECOexpert** is a module that provides advice on actions related to restoring a ecologically balanced environment after a chemical spill. Within the **SPILLexpert** system, the operating sequence of these modular expert systems can be altered based upon the priority of the target problem. Each module can be operated independently. The overall plan is to build a global expert system for chemical spill accidents in an incremental manner by assembling a number of modular expert systems into a defined structure (a bottom-up approach).

4.4 COMPONENTS IN THE EREXPERT PROGRAM

4.4.1 The knowledge base

ERexpert allows different knowledge representation formats to be used simultaneously. The overall knowledge base in this program contains both a Rulebase and a Factbase. The Rulebase is an ASCII file in which heuristic knowledge is coded as production rules that are associated with a number of goal variables. The Factbase comprises a number of subject-oriented tables in which factual information is stored as either text or numbers [17]. During execution, the heuristic module containing "IF condition THEN conclusion" rules can be used to offer the user advice based on the facts provided, while a search of the Factual database provides information specific to the case that can supplement the conclusion(s) reached during an inference process.

4.4.2 The inference engine

The inference engine, EAengine, a part of EAshell [26], supports three inference strategies, namely forward, backward, and mixed chaining Strategies of each of these three inference processes are described in Section 2.2 and will not repeated here.

4.4.3 The user interface

One of the major problems to be solved in the development of an expert system is to meet the requirements of an effective yet easy-to-use user interface that will allow the efficient use of the outcome systems. The object of the **ERexpert** system is to make planning the response to a spill faster and more accurate, so the design of the user interface is a critical feature of the program There are two layers of communication in **ERexpert**: (i) between the modular expert systems, and (ii) between the system user and the program. MS Windows© is a powerful operating system that provides communication functions through dynamic data exchange (DDE), object linking and embedding (OLE), dynamic linking libraries (DLL), and the multiple-document interface (MDI). A graphical user interface (GUI) for the **ERexpert** system was developed [39]. This interface was written using Microsoft Visual Basic© (ver 3.0).

4.5 KNOWLEDGE BASE DEVELOPMENT

Knowledge in most areas of specialization can be classified into two categories: factual knowledge that may be obtained from published sources, and heuristic knowledge that is used by human experts in manipulating and interpreting factual knowledge [40]. Factual knowledge, such as is held in theories and in mathematical algorithms, is usually carefully defined, clearly expressed, and well documented. One can acquire this type of knowledge through learning processes. However, when faced with a real-world problem, it is common to find that the problem cannot be solved solely using factual knowledge, therefore experience or heuristic knowledge enters the problem-solving arena. Unlike factual knowledge, heuristic knowledge often does not have clear-cut values and is hard to express or document precisely. One can only acquaint oneself with heuristics through practice, a process of applying the acquired factual knowledge. Expertise as practiced by an expert is a combination of both these knowledge categories.

An expert system requires more than factual knowledge before it can be applied to solve domain problems involving uncertainty. The factual information held in the knowledge base is essential in providing operational details and completing the knowledge representation for the problem domain, however, it is the heuristic knowledge in the knowledge base that extends the flexibility and applicability of a knowledge based system. Penninckx, *et al.* have argued that a database of information is an extension of a knowledge system because it guarantees the completeness of the information represented [41]. As the information source of an expert system, the knowledge base containing both heuristic and factual knowledge has to be compiled into a format that will allow the inference engine to match actual conditions with the coded knowledge in order to identify a conclusion. The knowledge base used by the **ERexpert** prototype encompasses two components: (i) a Factbase, which is a database of factual knowledge relevant to the problem domain, and (ii) a Rulebase, which uses the rule format to represent the heuristic knowledge. The inference engine used in this work manipulates the heuristic knowledge represented in rules from the Rulebase to interpret associated factual knowledge in the Factbase. The factual knowledge was obtained from published materials, and the domain expertise was extracted through interviews and discussion with field experts.

4.5.1 Structure of the Factbase and subsequent development

Certain types of information involved in the spill response belong to the factual knowledge category. For example, the phone numbers and addresses of local environmental agencies and industrial organizations, the assessment techniques, physico-chemical properties, containment equipment, etc., will not change with target chemical(s) or the geological location of an accident Also, a collection of countermeasures used in previous spills is an important source of information for reference under the new circumstances of the current spill Such information is referred to as static information and is organized into the Factbase to supplement facts necessary for planning response actions.

Structure

Microsoft ACCESS© was chosen as the database system for construction of the Factbase. The Factbase of **ERexpert** consists of a number of tables, each containing a different subject relevant to the problem domain Table 4 2 lists the contents of this Factbase. Three layers of fact tables exist in this database the primary tables, the secondary tables, and the auxiliary tables [42]. Primary tables are a set of fundamental tables that determine the coverage of the current Factbase and provide records of previous cases and the general response techniques. Secondary tables consist of a property section and a containment technique section. The property section is designed to supplement each data entry in the primary level with a set of physico-chemical properties as references for planning. The technical section consists of a group of tables that provide technical support for the remedial methods consistent with the target chemical(s) and the nature of the spill. The auxiliary tables provide the system user with general information and procedures necessary to handle a spill. The Microsoft ACCESS© database system offers powerful query and macro functions that a system developer can use to pre-define the relationships between different tables. Refer to Appendix 1 for more descriptions about the design of this database.

Table 4.2 Contents of the Factbase in the ERexpert module.

Primary Tables

1. Compound List
2. Case History Database
3. Containment Techniques

Secondary Tables

Property Tables:

1. Toxicity and First Aid
2. Fire and Explosion Data
3. Chemical Property Data
4. Physical Property Data

Technique Tables:

1. Adsorption Techniques
2. Neutralization Techniques
3. Precipitation Techniques
4. Chelation Techniques
5. Solvent Extraction Techniques
6. Reduction/Oxidation Techniques
7. Bio-remediation Techniques
8. Incineration Techniques

Auxiliary Tables

1. Phone Directory of Regional Response Centers
2. Shipping and Storage Requirements
3. Protective Equipment List
4. Disposal Requirements / Procedures

In Figure 4.4, the structural details of the Factbase used by **ERexpert** are fully described The diagram illustrates the contents of each table and how they are interrelated. In Figure 4.4, the "1" symbol indicates the "one" side of the relationship, and the "M" symbol indicates the "many" side of the relationship. The information stored in the fact tables is chained together through a group of identification numbers (ids). The *Comp ID* is an id number used to connect information in the property tables and the technical tables with the primary tables. Containment and remedial techniques are organized into the technical database by order of the environmental behavior of each chemi An index

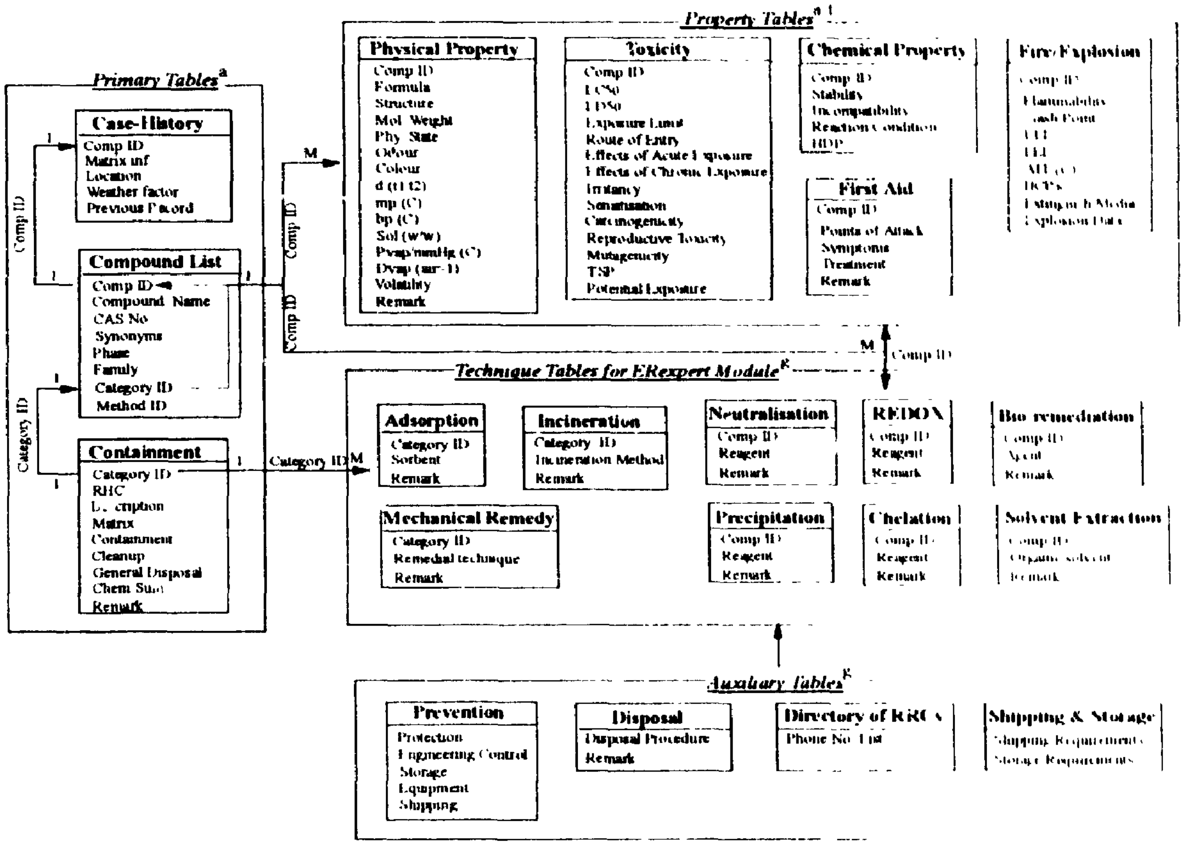


Figure 4.4 A diagram showing the structure of the Factbase of the ERexpert prototype.

The "1" symbol indicates the "one" side of the relationship. The "M" symbol indicates the "many" side of the relationship. Accordingly, information of different sources(a-g) is arranged into this subject- oriented structure Through embedded macro and query functions, the database is able to provide user specified information dynamically in various formats.

a. Lewis, R. J., Sr. Hazardous Chemicals Desk Reference, 2nd Ed.; VNR: New York, 1991
b. CRC Handbook of Chemistry and Physics, 75th Ed.; Lide, D R , Editor-in-Chief; CRC Press Ann Arbor, 1994/95.
c. Sittig, M. Handbook of Toxic and Hazardous Chemicals and Carcinogens, 3rd Ed , Noyes Publications: New Jersey, 1991; Vol. 1-2.
d. Kaye, S. Handbook of Emergency Toxicology. A Guide for the Identification, Diagnosis, and Treatment of Poisoning, 5th Ed.; Charles Thomas Publisher: Springfield, Illinois, 1988
e. The Merck Index. An Encyclopaedia of Chemicals, Drugs, and Biologicals, 11th Ed , Budavari, S., Ed.; Merck & Co., Inc.: New Jersey; 1989.
f. Amdur, M. O.; Doull, J.; Klaassen, C. Casarett and Doull's Toxicology. The Basic Science of Poisons, 4th Ed.; Pergamon Press: New York, 1991.
g. Hazardous Chemical Spill Cleanup, Pollution Technology Review No 59, Robinson, J S , Ed.; Noyes Data Corporation: New Jersey, 1979.

(*Category ID*) is assigned to each behavioral group to provide connections with the primary tables. The *Method ID* is reserved for use in the development of the ACmethod expert system module. Details of the Factbase used by the **ERexpert** module are given in Appendix 1. This subject-oriented structure helps to reduce the complexity in the development of an information database designed to manage elaborate information, and it also provides the necessary flexibility in both the database development stage for knowledge input, and the post-development stage when the database needs to be upgraded. Furthermore, this subject-oriented structure is more efficient during execution since it requires less memory to load only the relevant part of the database.

Development

The inventory of chemicals changes continuously. To develop a database that can handle spills, it is neither expeditious nor practicable to attempt to develop a conventional database, one that could cover the entire chemical family, and to include every potential remediation method against each hazardous chemical. It also cannot be overemphasized that the appropriate response to a chemical spill varies from case to case, it should not be assumed that countermeasures adopted in previous cases can be used without modification under the new circumstances. A practical approach to address this conflict is to develop a categorization method that divides hazardous chemicals into different groups based on their environmental behavior so that chemicals in the same category can be assessed and similar countermeasures used. Such a categorization leads to a knowledge-based system that can be deployed to deal with the majority of hazardous chemicals [43].

Table 4.3 lists a set of key physico-chemical properties that have been identified as criteria that can be used to categorize chemicals into different environmental behavioral groups. Using these criteria, chemicals were divided into four major categories and 17 sub-groups. A training set containing seventeen representative hazardous chemicals (RHCs) was identified so that each RHC represented a sub-group of compounds. Following the structure of the subject-oriented database, the Factbase was then developed. The "Compound List" table of this Factbase contained the RHCs and the set of ids assigned to them. The physico-chemical data for these RHCs were arranged into the property tables to construct the property section of the database. Table 4.4 summarizes the most commonly used containment and remedial techniques, and this information was mapped into the technical section of the Factbase [43]. Procedures for shipping, storage, disposal, protection, etc. were added to the supplementary tables to serve as the auxiliary section of the database. Queries and macros were developed to provide the necessary connections between individual tables. These functions were developed using the Microsoft ACCESS© Basic language, a procedural program supplied with Microsoft ACCESS©. The user interface that communicates between the program, the database, and the system user was developed using Microsoft Visual Basic (version 3.0). It should be mentioned that the database structure discussed above was refined several times during tests. A more detailed description of this database will be given in Section 4.7.1.

Table 4.3 Key physico-chemical properties used by ERexpert for compound categorization

Criteria	Principal behavioral group	Sub-group/Category ID
a. physical state b. density relative to air	1. Chemical that vaporizes (G1)	1A, 1B, 1C, 1D
c. density relative to water d. flammability	2. Chemical that floats (G2)	2A, 2B, 2C, 2D, 2E
e. water solubility f. volatility	3. Chemical that sinks (G3)	3A, 3B
g. acid/base (including those react with water to give acid or base)	4. Chemical that dissolves (G4)	4A, 4B, 4C, 4D, 4E, 4F
h. biodegradation i. salt containing heavy metal ion		

Table 4.4 Summary of containment and remedial techniques and specifications[a,b]

IN AIR		ON LAND (liquid or solid)				IN WATER (liquid or solid)					
Vapor		Solid		Liquid		Sinking (insoluble)		Floating		Soluble & Miscible	
Technique	*Specification*	*Technique*	*Specification*	*Technique*	*Specification*	*Technique*	*Specification*	*Technique*	*Specification*	*Technique*	*Specification*
Dispersion.	Very calm, sheltered areas	Self containing		Earthen dikes and trenches	Flat or sloped surface	Natural dikes & excavation	Where a natural barrier exists	Booms	Not much current	Sealed booms	Limited area, contain depth
Inert foam coverage	Low-lying vapors and calm areas			Foamed polyurethane	Hard, dry surface	Under water dikes and excavations	If bottom can be moved	Weirs	Calm water surface	Diversion of uncontaminated water	Flowing water, clear area needed
Cryogenic cooling.	Sheltered areas			Foamed concrete.	Flat ground, slow moving spill	Curtain barriers	Calm water low disperse	Pneumatic barriers	Shallow and calm water	Diversion of contaminated water	Flowing water, clear area needed
				Excavation.	Soft ground.					Gelling agent	Small volume
				Soil surface	Soft ground.					Water dispersion	
Mist knock down	Water-soluble or low-lying vapors	Shoveling & vacuuming.	Under normal weather conditions.	Correction of mechanic failure	Close valve, shut down pump, pudding to pipe.			Skimmer.	Calm water surface.	Adsorption	Most wide used method
Air dilution using fans or blower	Very calm & sheltered areas			Pump or vacuum collection into auxiliary tank or sump.		Pump.		Vacuum collector	Calm water surface	Neutralization, Ion exchange, Precipitation, Chelation, and Redox	Chemical specific methods
Cryogenic condensation..	Very calm & sheltered areas			Earth moving.	Removable land	Dredging	Bottom removable	Burning	Geographically right areas	Centrifuge separation, solvent extraction, gelation	Small removable water volume
Encapsulation	Low-lying & limited spill size.			Burial	Temp. mitigating measure with very limited application	Burial	Temp mitigating measure with very limited application			Biodegradation	Less toxic chemicals

a. Hazardous Chemical Spill Cleanup, Pollution Technology Review No. 59; Robinson, J. S., Ed.; Noyes Data Corporation: New Jersey, 1979.
b. Lewis, R. J., Sr. Hazardous Chemicals Desk Reference, 2nd Ed.; VNR: New York, 1991.

Operation

Figure 4.5 is a flow chart that illustrates the search procedures used with the knowledge base The complete search process comprises a database search process based on the match of key-words, a rule-directed inferential process, and a system update procedure (refer to the contents in Section 4.7). The key-word search process starts with the "Compound List" table to determine if there are valid database entries that could match the user-specified inquiry The search procedure has been developed to perform searches based on unique database entries (search strategies), such as *Compound Name*, *CAS No.*, as well as using synonyms of the target chemical.

If a search strategy is fulfilled, then a set of id numbers for the target is obtained. In this way, all facts about the target chemical that are stored in various factual tables in the Factbase are made available to the user. In addition, the search program uses the [Comp ID] parameter as the key word to search the "Case-History" database to see whether similar case(s) have bee reported previously. The "Case-History" database holds descriptions of previous spills and the corresponding countermeasures adopted This database can be upgraded to include the current plan under development using the upgrade procedure provided. In addition to the [Comp ID] parameter, three common questions are asked in order to restrict the number of possible hits during the search process. The questions to be answered are as follows: Q1. the matrix in which the accident happened, Q2. the type of location of the accident, and Q3. the weather. Predetermined answers to these questions are provided as menu options to further reduce the diversities of entries as following. For Q1, the options are: (a) watercourse, (b) loose surface, (c) paved surface, and (d) user specified. For Q2, the options are: (a) urban area, (b) rural area, (c) major transportation route, and (d) user specified. For Q3, the options are: (a) normal, (b) windy, (c) raining, (d) snowing, and (e) user specified. Ideally, the result(s) reached by the end of this second round of searching should

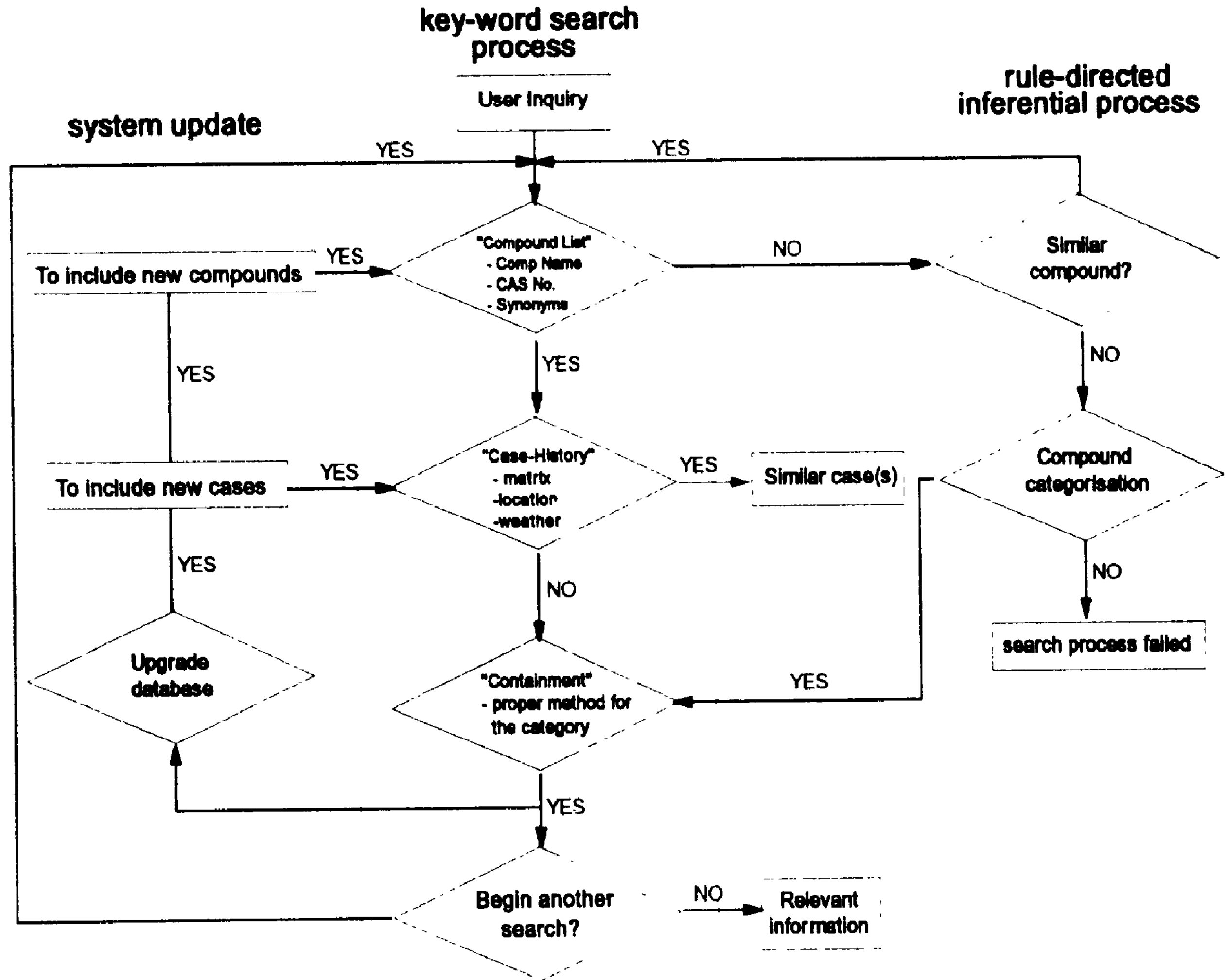


Figure 4.5 Operational diagram of the knowledge base.
The knowledge base is equipped with three different operation routines: (1) a key-word search process; (2) a rule-directed inferential process; and (3) a system upgrading process.

closely match the situation one is dealing with, and therefore, can be used as a guide to plan the response for the current spill. Practically, however, such occasions will be rare. Regardless of the outcome from the "Case-History" database search, if the primary search did not fail, one can continue to collect relevant information for particular compound(s) in the Factbase based on the connections provided by the id numbers.

Use of the system shows that this database structure is efficient in managing the massive amounts of factual information required in arriving at an appropriate response to a chemical spill. The search subroutines were found to be sufficiently fast to handle user requests and to output search results The test results also showed that this database could be easily upgraded to include additional data entries without modification of the internal structure.

4.5.2 Compilation of heuristic knowledge

Incomplete information or uncertainty in the knowledge is always the final dilemma when decisions have to be made. Human experts will use experience, or "rule of thumb", to resolve such predicaments. Edwards and Cooley argued that an expert system required more than factual knowledge to display expertise in a given domain [44]. They inferred the importance of heuristic knowledge in the problem-solving process. Kidd wrote that the aim of an expert system is not merely to capture a static representation of some knowledge domain but to simulate a particular problem-solving task carried out within that domain [45]

As described earlier, we developed the spill-handling database using a subject-oriented structure. The role of this database is to categorize hazardous chemicals and to provide characteristic information such that chemicals in the same group can be treated similarly. Nevertheless, real-world situations can be so dynamic that such a database may not be able to provide the diversity and resolve the complexities that arise during the problem-solving process. Thus, a knowledge-based system is necessary to extend the applicability and flexibility of the RHC database. In the **SPILLexpert** project, the knowledge base employs a production rule format to represent knowledge. In this section, we describe the results achieved in knowledge base development using the multi-layer KDM knowledge encoding methodology described in Chapter Two.

Knowledge compilation results

Applying the causal analysis operation to the contents in Table 4.3, we obtain Table 4.5 in which the criteria shown in Table 4.3 are listed as conditions and Category IDs are identified as conclusions. These results can be readily transferred into an empty knowledge matrix. Next, the logic connections are placed in the corresponding cells in the matrix. Figure 4.6 shows part of a completed KDM that combines the knowledge shown in Table 4.3 and the logical connections given in Table 4.5. More knowledge compilation results are described in Section 4.7.2 and in Appendix 2. In order for the inference engine to operate, the knowledge compiled in the KDM format must be converted into a set of conditional statements, known as production rules, in which the coded knowledge is represented as a series of IF condition(s) THEN conclusion(s) sentences without using specific artificial intelligence language. Consequently, the containment and remedial countermeasures listed in Table 4.4 and the information stored in Factbase can be selectively accessed to aid in planning response to a chemical spill when the Category ID has been identified.

Major components in the KBF file used by the **ERexpert** program are described in Table 4.6, and include: (i) the goal section containing a list of goal variables that will be used to assign the result(s) from the inferential process; (ii) the production rule section, which is the main section where heuristic knowledge is stored in IF...AND/OR...THEN... statements; and (iii) the user query section providing the inference engine with pre-defined questions. In an inference process, the inference engine will only ask those questions related to the facts that the user has entered and present options for identification. The current knowledge base of the **ERexpert** prototype is ready for experimental testing, we expect that the 'test with cases' cycle will introduce new conditions and conclusions into the knowledge base as the system is adjusted to become more practical and applicable for use in real situations.

Table 4.5 Causal analysis of the knowledge used for chemical categorization

1. Heavier than air, flammable vapors	{Group 1A chemicals: [Vapor state, Vapor density greater than air, Flammable]}
2. Heavier than air, nonflammable vapors	{Group 1B chemicals: [Vapor state, Vapor density greater than air, Non-flammable]}
3. Lighter than air, nonflammable vapors	{Group 1C chemicals: [Vapor state, Vapor density less than air, Non-flammable]}
4. Lighter than air, flammable vapors	{Group 1D chemicals: [Vapor state, Vapor density less than air, Flammable]}
5. Floating solids	{Group 2A chemicals: [Solid state, Density less than water, Water insoluble]}
6. Floating, flammable liquids	{Group 2B chemicals: [Liquid state, Density less than water, Flammable, Water insoluble]}
7. Floating nonflammable liquids	{Group 2C chemicals: [Liquid state, Density less than water, Nonflammable, Water insoluble]}
8. Floating, spreading liquids	{Group 2D chemicals: [Liquid state, Density less than water, Water insoluble, Spreading on water surface]}
9. Floating, non-spreading liquids	{Group 2E chemicals: [Liquid state, Density less than water, Water insoluble, Non-spreading on water surface]}
10. Sinking solids	{Group 3A chemicals: [Solid state, Density greater than water, Water insoluble]}
11. Sinking liquids	{Group 3B chemicals: [Liquid state, Density greater than water, Water insoluble]}
12. Dissolving, acidic chemicals	{Group 4A chemicals: [Water soluble, Acidic chemical, React with H_2O]}
13. Dissolving basic chemicals	{Group 4B chemicals: [Water soluble, Basic chemical/React with H_2O]}
14. Dissolving salts with heavy metal ions	{Group 4C chemicals: [Water soluble, Salt with heavy metal ion]}
15. Dissolving salts without heavy metal ions	{Group 4D chemicals: [Water soluble, Salt without heavy metal ion]}
16. Biodegradable chemicals	{Group 4E chemicals: [Water soluble, Biodegradable]}
17. Non-biodegradable chemicals	{Group 4F chemicals: [Water soluble, Non-biodegradable]}

Conditions \ Conclusions	Vapor chemicals				Insoluble chemicals							Soluble chemicals							
	1. Low lying, flammable vapors (1A)	2. Low lying, nonflammable vapors (1B)	3. Lighter than air, nonflammable vapors (1C)	4. Lighter than air, flammable vapours (1D)	5. Floating solids (2A)	6. Floating, flammable, volatile liquids (2B)	7. Floating, nonflammable, nonvolatile liquids (2C)	8. Floating, volatile, nonflammable liquids (2D)	9. Floating, nonvolatile, flammable liquids (2E)	10. Sinking solids (3A)	11. Sinking liquids (3B)	12. Acidic chemicals that dissolves in water (4A)	13.Chemicals that dissolve & react with water to give acid (4A)	14. Basic chemicals that dissolve in water (4B)	15. Chemicals that dissolve & react with water to give base (4B)	16. Water soluble salts containing heavy metal ions (4C)	17. Water soluble salts containing no heavy metal ions (4D)	18. Biodegradable chemicals (4E)	19 Non-biodegradable chemicals (4F)
1. Vapor state	T	T	T	T	F					F									
2. Vapor density greater than air	T	T	F	F															
3. Flammable	T	F	F	T		T	F	F	T										
4. Water soluble					F	F	F	F	F	F	F	T	T	T	T	T	T	T	T
5. Liquid state					F	T	T	T	T	F	T								
6. Volatile chemical					F	T	F	T	F										
7 Density less than water					T	T	T	T	T	F	F								
8. Acidic chemical												T	F	F	F				
9. Containing heavy metal ions																T	F		
10. Biodegradable																		T	F
11. React with water												F	T	F	T				

Figure 4.6 A sample knowledge domain matrix (KDM).
The knowledge encoded in the KDM representing the causal analysis of the knowledge employed for chemical categorization. Production rules can be readily derived from this knowledge matrix.

Table 4.6 Description of the Goal variables in the KBF file used by ERexpert

Goal Section:	Description:
1. Chemical Categorization (CategoryID)	used to assist the user in assignment of a proper Category ID for a target chemical if it is not included previously in the Factbase.
2. Compound Similarity Comparison (CompID)	used when the target compound is not included in the 'Compound List' table. This section helps the user to decide whether the target compound might resemble a 'similar' compound in the database, based on similarities among a set of physico-chemical ctiteria
3. Containment & Remedial Technique (ContTech)	used to assign proper containment and/or remedy methods for the target compound once this chemical's Category ID is decided.
4. Treatment Site Selection (TrmtSite)	used after an accident has been contained, this section helps the user to select a proper treatment site for further treatment or disposal based on the facilities, supplies and the geological location of the spill.

The rule-directed inferential process is shown in Figure 4.5. In this process, EAengine employs the knowledge stored in the Rulebase together with the supplementary information given by the Factbase to provide remediation advice following a spill. EAengine supports three inference strategies such as described in Section 2.2.

4.6 MODULE DESIGN

4.6.1 The SPILLexpert frame

The primary objective of this step was to design a structure that could enable a separate yet correlated development of a series of knowledge-based modules to solve problems of individual subdomain in a complex field of study Emphasis was paid to the feasibility and effectiveness of the integration of the independently-developed modules into this general structure As shown in Figure 4.7, the **SPILLexpert** frame was developed for response to chemical spill accidents in which the technical elements involved were incorporated into a

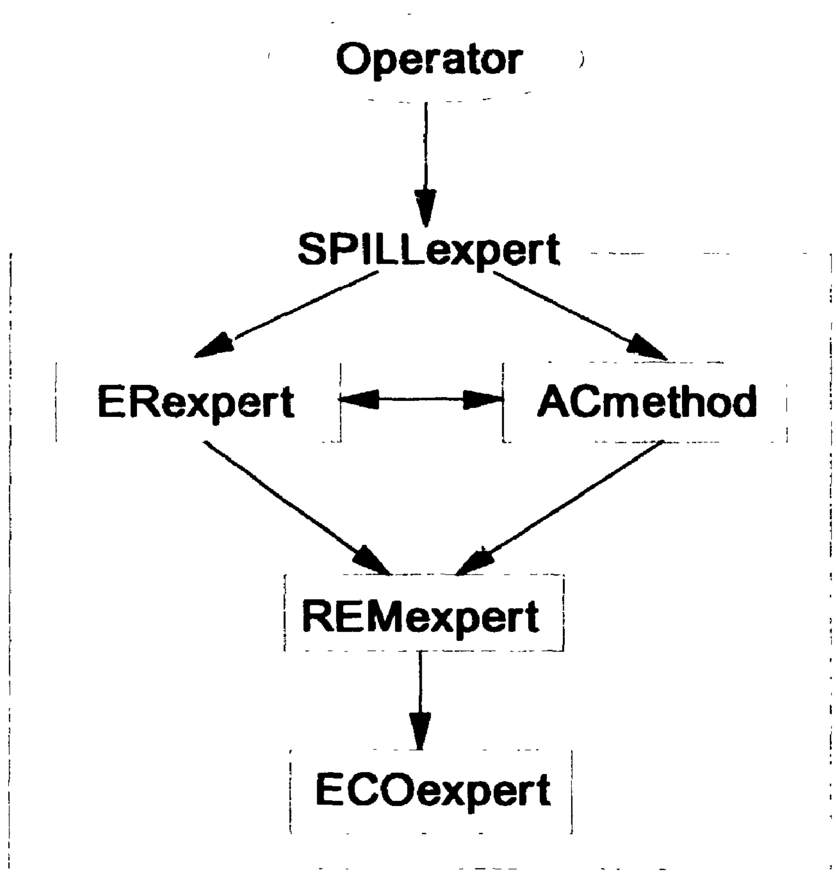


Figure 4.7 The SPILLexpert frame.
A set of interconnected modular expert systems designed so that ڊach focuses on a single subdomain of the overall problem domain.

number of modular expert systems refer Section 4.3 for the descriptions of each modular expert systems.

4.6.2 The ERexpert module

As a module in the **SPILLexpert** frame, **ERexpert** emphasizes the sub-domain of emergency response to a chemical spill that uses both production rules and a database of facts to represent the expertise of the domain experts. Figure 4.8 shows the design adopted for the development of the **ERexpert** module. The user interface section encompasses an I/O program that provides the functions of evidence inquiry (user input), explanation/display (system output), and the routines used for database access and for inference processes. The knowledge base section consists of both a Factbase and a Rulebase that can be used by the program interchangeably. The program first uses the static knowledge of the Factbase in an effort to solve a problem. If the facts about a particular spill are not sufficient, or exceed the limitations of the Factbase, the system then considers this problem to be an incomplete case and will launch the inference process to try and solve the uncertainties using the conceptual knowledge in the Rulebase. In **ERexpert**, both the database search process and the inductive procedures are computer subroutines embodied in the user interface that are controlled by the user.

4.7 IMPLEMENTATION

4.7.1 Implementation of the database structure

The factual database integrated into the **ERexpert** module, called Factbase, was developed using the MS ACCESS© database system (Appendix 1). This Factbase consists a set of internally connected, subject- oriented fact tables. The current Factbase of **ERexpert** contains 17 characteristic chemicals, called representative hazardous chemicals (RHCs), where each one represents one of the 17 sub-groups of chemicals that exhibit the behavioral categories

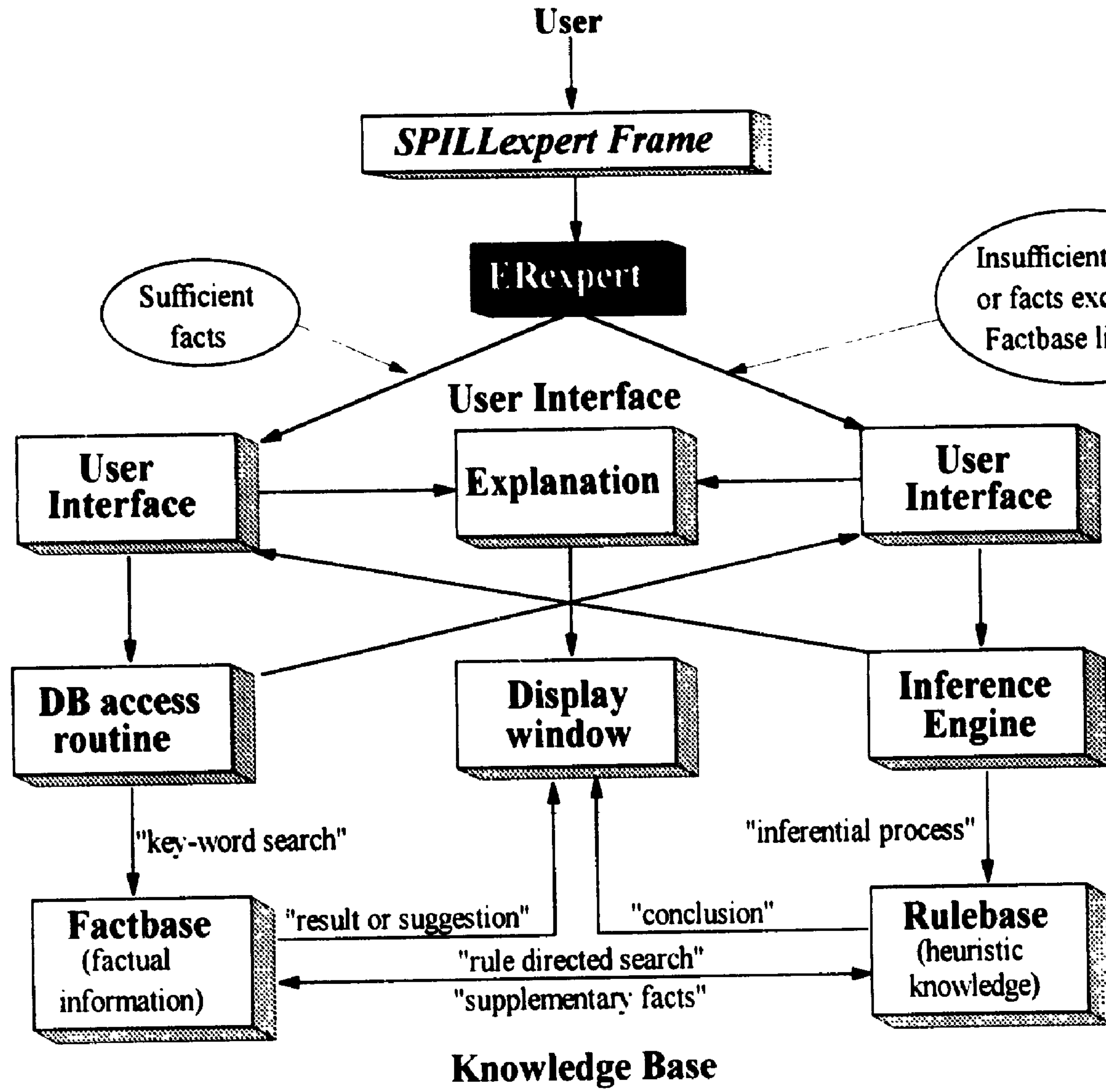


Figure 4.8 The design for the ERexpert program.
The diagram shows the interactions between each of the components and how information flows inside the program.

identified by the system [17]. These RHCs are used to mirror the whole community of hazardous chemicals for which **ERexpert** is designed to respond RHCs are stored in the "Compound List" table of the Factbase, such as illustrated in Figure 4.9. In this table, a group of pre-defined indices are included in association with the chemicals. These indices, called identification numbers (IDs), were employed to realize the connections among different factual tables. The internal connections between the tables are described in the diagram shown in Figure 4.10. According to this structure, relevant information for the chemicals listed in the "Compound List" table are mapped into separate subjective tables. The Factbase of **ERexpert** is furnished with query functions developed using the MS ACCESS BASIC© language. In conjunction with the ID numbers, these query functions provide accessibility to the factual data stored elsewhere in the Factbase. On execution, the database module runs as a background task. The database receives commands through the user interface to perform the required operations and returns the results to the front-end Additional factual information can be added fairly easily into the database according to the design, and, as is to be expected, the more the facts, the better the performance. However, the principal information set of the RHCs provide this database module with sufficient flexibility to cover a range of hazardous chemicals if the target chemical can be categorized into one of the 17 behavioral groups. Production rules designed to identify the chemical categorization are integrated into **ERexpert** and can be used to help with the identification if users have difficulty in determining the identity of a target compound.

As shown in Figure 4.10, in addition to the information supplied by the subject-oriented tables, the **ERexpert** Factbase is also equipped with a "Case-History" sub-database in which individual cases of responses to spills have been recorded and are connected to the main database through the <u>Compound ID</u> index. This "Case-History" sub-database serves two primary purposes. First, the records of previous responses to similar spills provides a

Comp ID	Category ID	CAS No.	Comp Name	Phase	Family	Synonyms	Method ID
5	1A, 2B	71-43-2	Benzene	VAPOR, LIQUID	Aromatic compound	Benzol, Phenyl hydride, Coal naphtha, Benxole, Cyclohexatriene	Reserved
36	1A; 2B	110-54-3	Hexane	VAPOR; LIQUID	Aliphatic hydrocarbon	None	Reserved
65	1B	7782-50-5	Chlorine	VAPOR	Halogen	Bertholite	Reserved
84	2A; 4E	108-95-2	Phenol	VAPOR, SOLID	Aromatic alcohol	Carbolic acid; Phenic acid; Phenylic acid; Phenyl hydrate; Hydroxybenzene, Monohydroxybenzene	Reserved
93	1C; 4B	7664-41-7	Ammonia, anhydrous	VAPOR, LIQUID	Ammonia	Ammonia	Reserved
98	1D	74-82-8	Methane	VAPOR	Aliphatic hydrocarbon	Mash gas; Methyl hydride	Reserved
144	3A	309-00-2	Aldrin	SOLID	Halogenated aromatic compound	1,2,3,4,10,10-Hexachloro-1,4,4a,5,8,8a-hexahydro-1,4,5,8-dimethano naphthalene	Reserved
264	3B; 4A	7726-95-6	Bromine	LIQUID	Halogen	None	Reserved
290	4A	7664-93-9	Sulfuric acid	LIQUID	Mineral acid	Oil of vitriol; Spirit of sulfur; Hydrogen sulfate. Sulfuric acid (fuming) is known as oleum.	Reserved
299	4B	1310-73-2	Sodium hydroxide	SOLID	Alkali hydroxide	Caustic soda; Caustic alkali, Caustic flake; Sodium hydrate; Soda lye; White caustic	Reserved
353	2C	107-21-1	Ethylene glycol	LIQUID	Alcohol	Glycol; Ethylene alcohol; Glycol alcohol; 1,2-Ethanediol	Reserved
400	3B	67-68-5	Dimethyl sulfoxide	LIQUID	NA	Methyl sulfoxide; DMSO; Gamasol 90	Reserved
401	4C	7761-88-8	Silver nitrate	SOLID	Salt	Lunar caustic; Silbernitrat	Reserved
402	4D	6484-52-2	Ammonium nitrate	SOLID	Salt	None	Reserved
403	2E	124-18-5	Decane	LIQUID	Aliphatic hydrocarbon	None	Reserved

Figure 4. 9 The "Compound List" table of the Factbase in ERexpert.

As part of the Factbase used in the ERexpert program, this table contains 17 characteristic chemicals (RHCs) each of which represents a subclass of chemicals in one of the 17 behavioural categories available to ERexpert. The RHCs are used to mirror the whole community of hazardous chemicals for which ERexpert can respond.

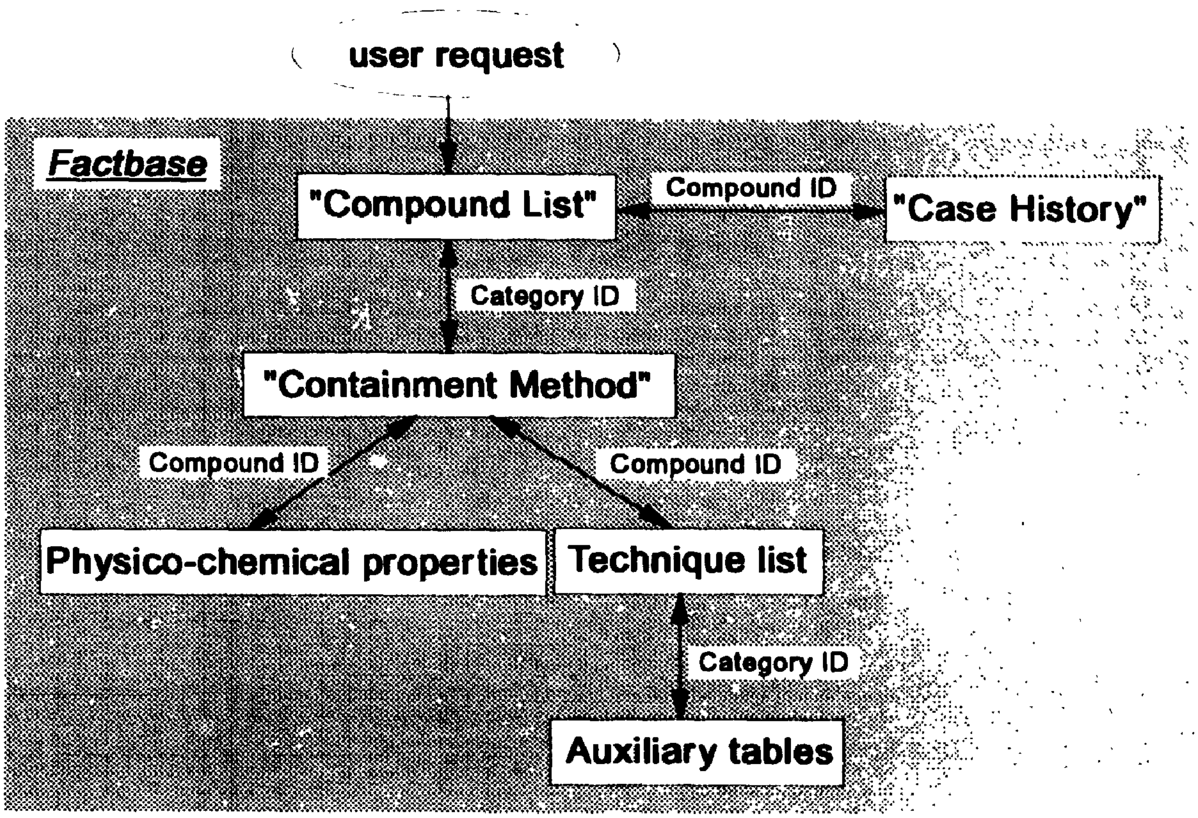


Figure 4.10 The structure of the Factbase.
This database is designed to encompass the factual information used in response to chemical spill accidents. The Factbase consists a set of internally connected, subject-oriented fact tables. A group of pre-defined indices are employed to enable the connections between the different fact tables.

valuable reference when planning response actions for the current situation. Second, this database provides an archive medium for effective documentation of the actions taken and reasons for those actions for each spill. Obviously, such a systematic recording of the actions taken in response to a spill will greatly aid in the sharing of experience. We anticipate that with the use of the **ERexpert** system, this sub-database will grow larger and play an increasingly useful role in the process of planning responses.

4.7.2 Implementation of the knowledge encoding results

The multi-level KDM structure has been implemented in the **SPILLexpert** project for the first time to represent the complex layers of knowledge involved in planning the appropriate response actions following a chemical spill emergency. Earlier in this chapter, we described a KDM for the knowledge sub-domain used in the classification of hazardous chemicals into different environmental behavioral categories. The contents of the second KDM layer is illustrated in Figure 4.11, which forms part of the complete KDM currently employed by the **ERexpert** system. The knowledge set encoded in this KDM is the knowledge sub-domain used in selection of proper containment techniques and/or primary remedial methods. As shown in Figure 4.11, a number of these conditions (denoted by asterisks) are conclusions from the previous round of inference. At the present stage, the heuristic knowledge in the **ERexpert** module has been encoded into a three-layer knowledge matrix, which includes the knowledge sub-areas of (i) compound categorization (Figure 4.6), (ii) containment and remedial methods selection (Figure 4.11), and (iii) treatment site selection (Appendix 2). Depending on the scope of a domain problem, these layers of knowledge can be applied separately or in conjunction with each other in the problem-solving process.

The coded heuristic knowledge in a multi-layer KDM structure needs to be reformatted before EAengine can perform the required inductive processes. In

Conditions / Conclusions	1 Correct mechanic failure, close valve, clog leaks, shut down pump, decide CateID, run KB again.	2.No spark, inert foam cover, mist knock down	3 No spark, cryogenic condensa., air dilution	4.Avoid spark Encapsulation.	5.Inert foam cover., mist knock down	6.Cryogenic condensat., air dilution.	7.Encapsulation.	8.Solid chemicals usually self-contained Shoveling/vacuuming recover	9 Fludized sol treated as liq , prevent run off into water/sewage, assign liq Cate ID to target	10 Treat as 2A/3A on dry surface	11 Treat as fluidized 2A/3A chem	12 No spark Inert foam/plastic sheet, earth dike. excavat. trench. pump/vacuum in auxi tank/sump, remove bottom, to paved surf /sealant slow penetrat. temp burial
1.Mechanical failure	T											
2.Category 1A chemicals*		T	T	T								
3.Category 1B chemicals*					T	T	T					
4.Category 1C chemicals*						t						
5.Category 1D chemicals			t									
6.Sheltered area			T			T						
7.Calm area (air/water suf.)		T	T	T	T	T	T					
8.Limited spill size				T			T					
9.Category 2A chemicals*								T	T			
10.Category 3A chemicals*								t	t			
11.Spill happened on land								T	T	T	T	T
12.Dry ground surface								T	F	T	F	
13.Category 2B chemicals*												T
14.Category 2C chemicals*												
15.Category 2D chemicals*												
16.Category 2E chemicals*												
17.Category 3B chemicals*												
18.Flat surface												
19.Soft surface/river bottom												T
20.Flammable												
21.Volatile												
22.Liquid state										F	F	
23.Spill in watercourse												
24.Category 4A chemicals*										T	T	
25.Category 4B chemicals*										t	t	
26.Category 4C chemicals*										t	t	
27.Category 4D chemicals*										t	t	
28.Category 4E chemicals*										t	t	
29.Category 4F chemicals*										t	t	
30.Natural barrier exists												
31.Geographically feasible												
32.Small removable volume												
33.Clear area												

														T					T			T											13.Earthen dike, trench, excavat., pump/vacuum to auxi. tank/sump, remove bottom, to paved surf /sealant slow penetrat , temp. burial
														T				T				T											14 Inert foam/plastic sheet, excavat., earthen dike/trench contain, pump/vacuum in auxi. tank/sump, remove bottom, to paved surf. /sealant slow penetrat., temp. burial
														T			T					T											15.No spark contact. Earthen dike/trench/ excavat., pump/vacuum in auxi. tank/sump, remove bottom, to paved surf./sealant slow penetrat., temp. burial for later treatment.
														F	T					T		T											16.No spark. Inert foam/plastic sheet, foamed polyurethane, concrete dike contain, pump/ vacuum in aux. tank/sump, to paved surf./ sealant, temp. burial for later treatment.
														F	T				T			T											17.Foamed polyurethane/concrete dike, pump /vacuum in aux. tank/sump, to paved surf. /sealant, temp. burial. for later treatment.
														F	T			T				T											18.Inert foam/plastic sheet, foamed polyurethane, concrete dike contain, pump/vacuum in auxi. tank/sump, to paved surf./sealant, temp. burial.
														F	T		T					T											19.No spark contact, foamed polyurethane, con -crete dike contain, pump/vacuum auxi. tank /sump, to paved surf./sealant, temp. burial.
												T	T			T						T											20.Treat as Cate 2B chemicals.
												F	F			T						T											21.Treat as Cate 2C chemicals.
												T	F			T						T											22.Treat a', Cate 2D chemicals.
												F	T			T						T											23 Treat as Cate 2E chemicals.
			T							T						T							t										24.Natural dike/excavation, pump recover in auxi tanks/temp. burial.

										T				T		T							t										25.Under water dikes, excavations, dredging, temp. burial.
										T						T							t			T							26.Curtain barriers, pump recover.
										T							t			T						T							27.No spark, booms/weirs/pneumatic barriers contain. skimmer/vacuum into auxi. tank.
										T								t	T					t		T							28.Booms/weirs/pneumatic barriers. Skimmers/ vacuum into auxi. tank.
		T								T							t			T													29.No effective method, burn.
										T								t	T					t									30.No effective contain/cleanup meth. for float. chemicals, water dilut./biodegrad.
	T									T						t	t	t	t	T			t	t									31.Removable insoluble chem , pump to auxi. tank/sump, centrifuge separat./solvent extract.
				t	t	t	t	t	T		T	T	T									T											32.Treat as Cate 2B chemicals.
				t	t	t	t	t	T		T	F	F									T											33.Treat as Cate 2C chemicals.
				t	t	t	t	t	T		T	T	F									T											34.Treat as Cate 2D chemicals
				t	t	t	t	t	T		T	F	T									T											35.Treat as Cate 2E chemicals.
				t	t	t	t	t	T	T																							36.No effect. meth for dissolving chemical in large, open water, water dilut./biodegradat.
	T			t	t	t	t	t	T	T																							37 Removable vol soluble chem.(4A-4F), pump to auxi tank/sump, solvent extract /gelation.
									T	T															T	T							38 Sealed boom in limited, calm water, neutralizat. except HCN, also ion exchange/adsorpt
T									T	T															T								39 Limited size/flow water. divers . neutralizat except HCN, adsorpt . ion exchange
								T		T															T	T							40 Sealed booms in limited, calm water, CO2, NaHCO3 neutralizat , adsorp /ion exchange
T								T		T																							41 Limited size/flow water, divers . neutralizat , adsop (activated C), ion ex (process resin)
							T			T															T	T							42 Sealed booms cont , precipit , adsorp ion ex

T							T			T															T								43 Limited size/flow water, diverse, precipt., adsorp., ion exchange.
						T				T															T	T							44.Sealed booms contain, oxidat. CN salt, adsorpt ion ex other salts in the group
T						T				T															T								45 Divers., oxidat (CN), adsorp./ion ex(others)
					T					T															T	T							46.Sealed booms contain, adsorpt /ion ex
T					T					T															T								47.Limit size/flow water, divers., adsorpt., ion ex
				T						T															T	T							48.Sealed booms contain, adsorpt., ion ex
T				T						T															T								49.Limit size/flow water, divers., adsorpt , ion ex

Figure 4.11 The Knowledge Domain Matrix (KDM) of the second knowledge layer in the ERexpert program.

As part of the entire KDM (refer to Figure 4.6 and Appendix 2 for the rest of the KDMs) used in ERexpert, the knowledge encoded in the KDM is the knowledge subdomain regarding the selection of containment techniques and/or primary remedial methods for chemical spills. Those conditions denoted by asterisks are conclusions from a previous round of inference. The heuristics in ERexpert have been encoded into a three-layer knowledge matrix including: (i) compound categorization; (ii) containment/remedial method selection; and (iii) treatment site selection. Depending on the scope of a target problem, these layers of knowledge can be applied separately or in conjunction with each other. "t"s represent the "parallel" cases in the KDM that are parallel to the "T" in the above adjacent cells.

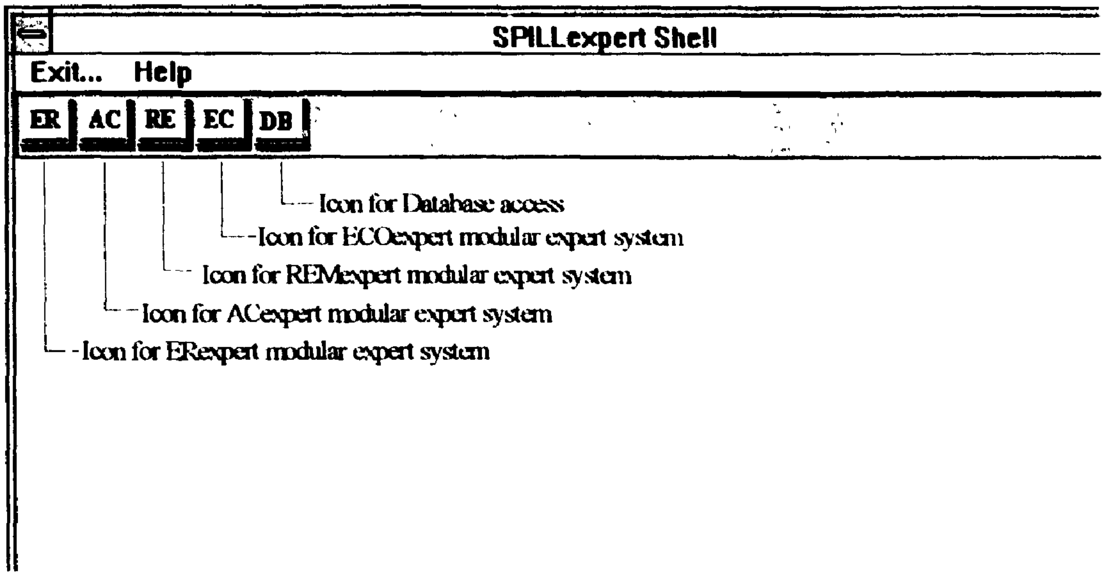
this study, the results obtained from knowledge matrices are converted into a set of conditional statements, known as production rules, in which the heuristics are represented in the form of a series of IF conditions) THEN conclusions) statements. A tool kit (Rule-Editor) has been developed in our laboratory to convert a filled KDM into a knowledge file of rules (KBF) for use by EAengine [26]. A KBF file consists of three major sections: a goal section, a logical expression section, and a user query section. The goal section provides information about goals used for backward chaining and contains a set of goal-related variables. The logical expression part is the main section in which heuristics are represented in *IF...[AND/OR]... THEN...* statements, an explanation (*EXP*) statement is also given for conclusions of each rule. Finally, the user query section asks questions related to the facts and provides options as answers for identification. An example KBF file and a brief explanation are given in Table 4.7, in which CateID, ContTech, and TrmtSite are goal variables that represent the knowledge sub-domains in the KBF file. CateID represents the knowledge sub-domain used for categorization of chemicals based upon their environmental behavior, ContTech is a knowledge sub-domain that deals with containment and remedial actions, and TrmtSite describes the heuristic knowledge used in selection of the treatment site. A completed version of knowledge files (KBF) used in the SPILLexpert project is presented in Appendix 3 which contains 99 production rules with 66 variables. It is important to note that when both AND and OR are used in a rule, OR takes precedence over AND as if parentheses surround expressions connected by OR. This rule (named 29) is applicable if the spill happened on an unpaved surface (highway side, farm land, etc.), and solid chemicals were involved that either float or sink if they encounter water. The conclusion suggests that if the ground is wet, the target chemical(s) should be treated as liquid chemicals and the major work should be directed towards preventing run off into the watercourse or sewage system.

Table 4.7 A rule file (KBF) example and brief explanations of the variables.

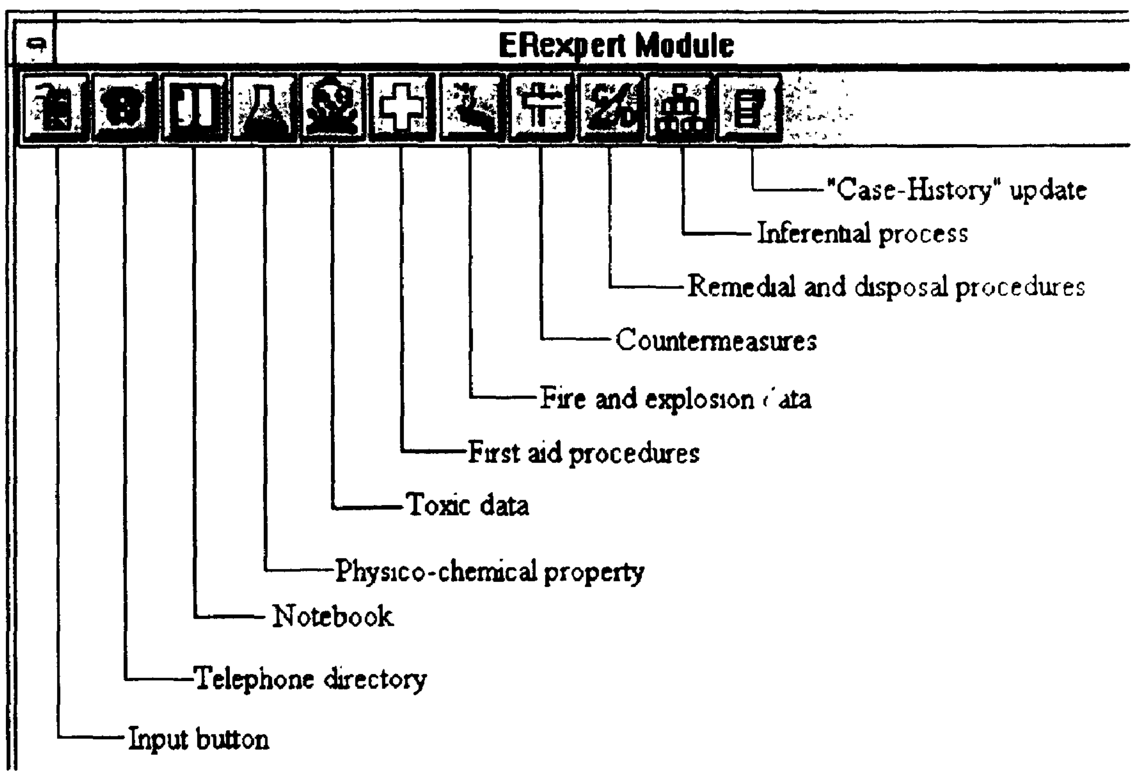
KBF Example	**Explanation**
Goal Section **FIND** CateID, ContTech, TrmtSite; :	Starts with ***FIND***, ends with ";". Goal variable **CateID** is for the assignment of a categorization ID to a compound; **ContTech** for selection of containment techniques; and **TrmtSite** for selection of treatment site.
Logic Expression Section RULE 29 **IF** Variable31=YES **AND** Variable32=NO **AND** Variable29=YES **OR** Variable30=YES	This block begins with keyword ***IF***, the premises are connected by ***AND***s or ***OR***s, and the pointer ";" ends the block.
THEN ContTech=As liq./Prevent run-off/Assg. liq.ID; **EXP** "Fluidized solids shall be treated as liq., major concern is to prevent from running off into watercourse or sewage. Response methods may be obtained by assigning a proper liquid category ID to the target." :	Corresponding content of each variable is given by the ASK sentence in the Query Section. The content after Keyword ***THEN*** is the conclusion part of a rule block, and a full explanation of the result is given in a ***EXP*** block immediately after that rule block.
User Query Section **ASK** Variable31 "Spill happened on land?" **OPTION** YES, NO; **ASK** Variable32 "Dry ground surface?" **OPTION** YES, NO; **ASK** Variable29 "Category 2A compound?" **OPTION** YES, NO; **ASK** Variable30 "Category 3A compound?" **OPTION** YES, NO; :	The ***ASK*** keyword begins a query block while a ";" ends it. ***OPTION*** clause provides available answers to a particular question.

4.7.3 Development of the ERexpert Module

Figure 4.12 shows the graphical user interface developed for the **SPILLexpert** project. This interface was written in the MS Visual Basic language within the MS Windows© 3.1 environment. A user with little computer knowledge can easily become familiar with the operation of each of the graphical features, such as the content of the menus and the intuitive icons. Part A in Figure 4.12 shows the main window of the **SPILLexpert** program in which five icons connect to five stand-alone modular expert systems embodied in the main frame. Clicking on the icon ER activates the **ERexpert** module. In part B of Figure 4.12, the first screen of **ERexpert** is displayed in which an icon set is illustrated and briefly explained. This icon set provides the user with full access to the embedded functions in **ERexpert**. The operational sequence of **ERexpert**



(A)



(B)

Figure 4.12 The graphical user interface.
Part A shows the main window of the SPILLexpert program in which five icons connect to five stand-alone modular expert systems employed in this main frame. Part B is the first screen of ERexpert in which an icon set is illustrated and briefly explained. This icon set provides the user with a full access to the embedded functions.

is described in the flowchart given in Figure 4.13. The sequence starts with a database search for the target compound that the user provided. Based on the result of this process, different routines can be initiated that will deduce information about the chemical.

To run the **ERexpert** module, the user needs first to provide the system with a set of primary data about the spill. The template shown in Figure 4.14A is the input dialogue box. In the next step, **ERexpert** employs this information set to search the Factbase. If the target compound can be found in the database, the search results are assigned to a set of variables and **ERexpert** informs the user through the "System Reminder", a text bar at the bottom of the window (B), that a successful search process has been completed. By clicking on appropriate icons, the system is prompted to show corresponding information about the target chemical in a display window (B). Figure 4.14B shows a report sheet of the physico-chemical properties of the target compound (i.e. benzene). Customized report sheets were designed as the output medium to show the results of Factbase search in a more organized and easy to read format.

On the other hand, if the target compound is not currently included in the Factbase, **ERexpert** returns a failure message for the Factbase search process and prompts the user to start an inference process. As shown in Figure 4.13, one of the three inference strategies incorporated into the program, mixed chaining, forward chaining, and backward chaining, needs to be selected at the beginning of each inference process. In practice the mixed chaining mode is the most commonly used inference strategy.

The template shown in Figure 4.14C is the window designated for the mixed chaining mode. In this mode, a set of questions will be prompted by the system. The order of the questions presented by **ERexpert** is dynamic in that it is decided by statistical analysis of the knowledge matrix based on the answers to the previously-asked questions. Once a unique set of questions has been

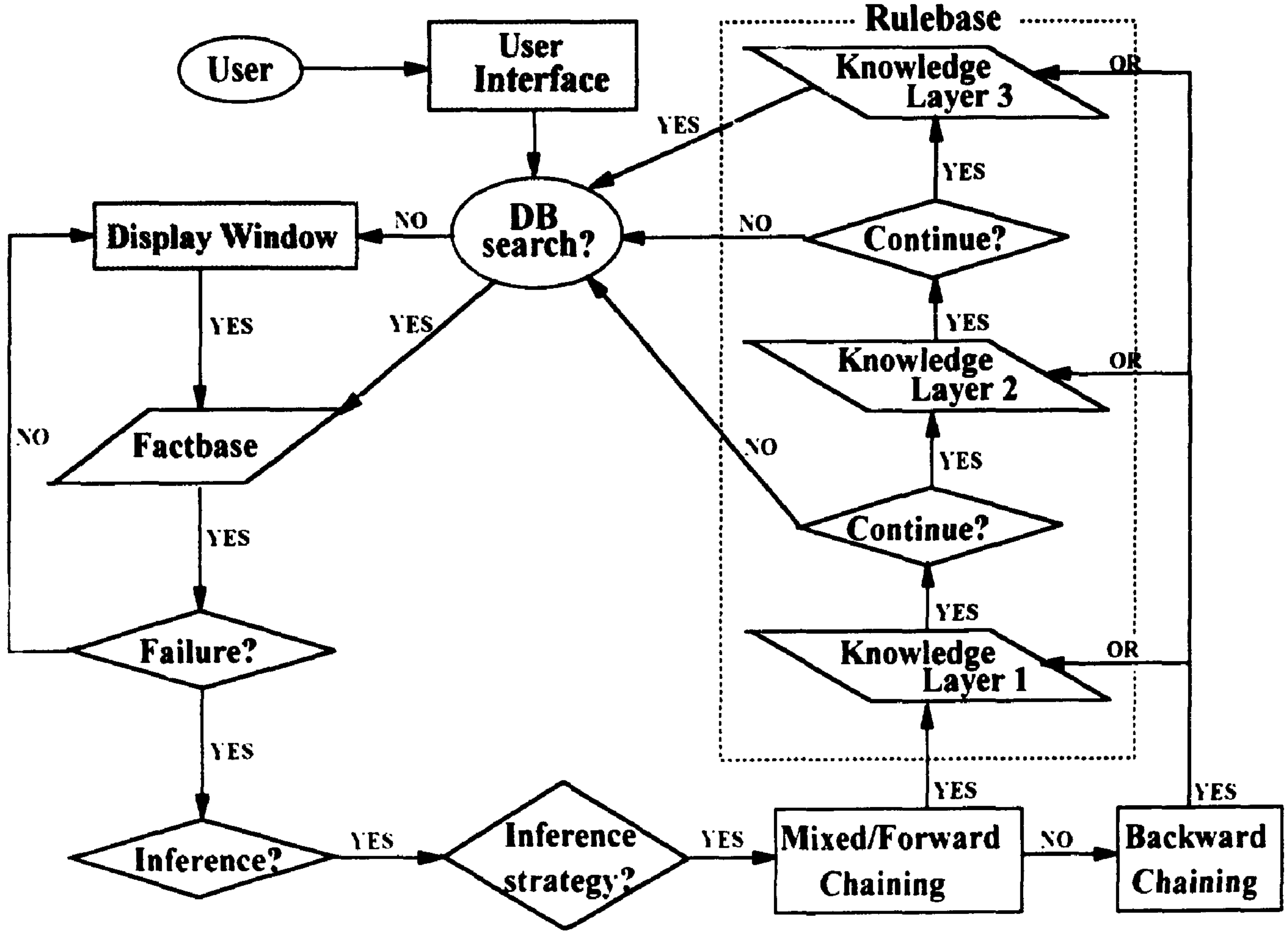


Figure 4.13 The operational sequence of the ERexpert prototype.

First, a database search process, that is followed by inference processes. Three inference strategies, forward, backward, and mixed chaining, are incorporated into the program, the chaining mode needs to be selected at the beginning of each inference process.

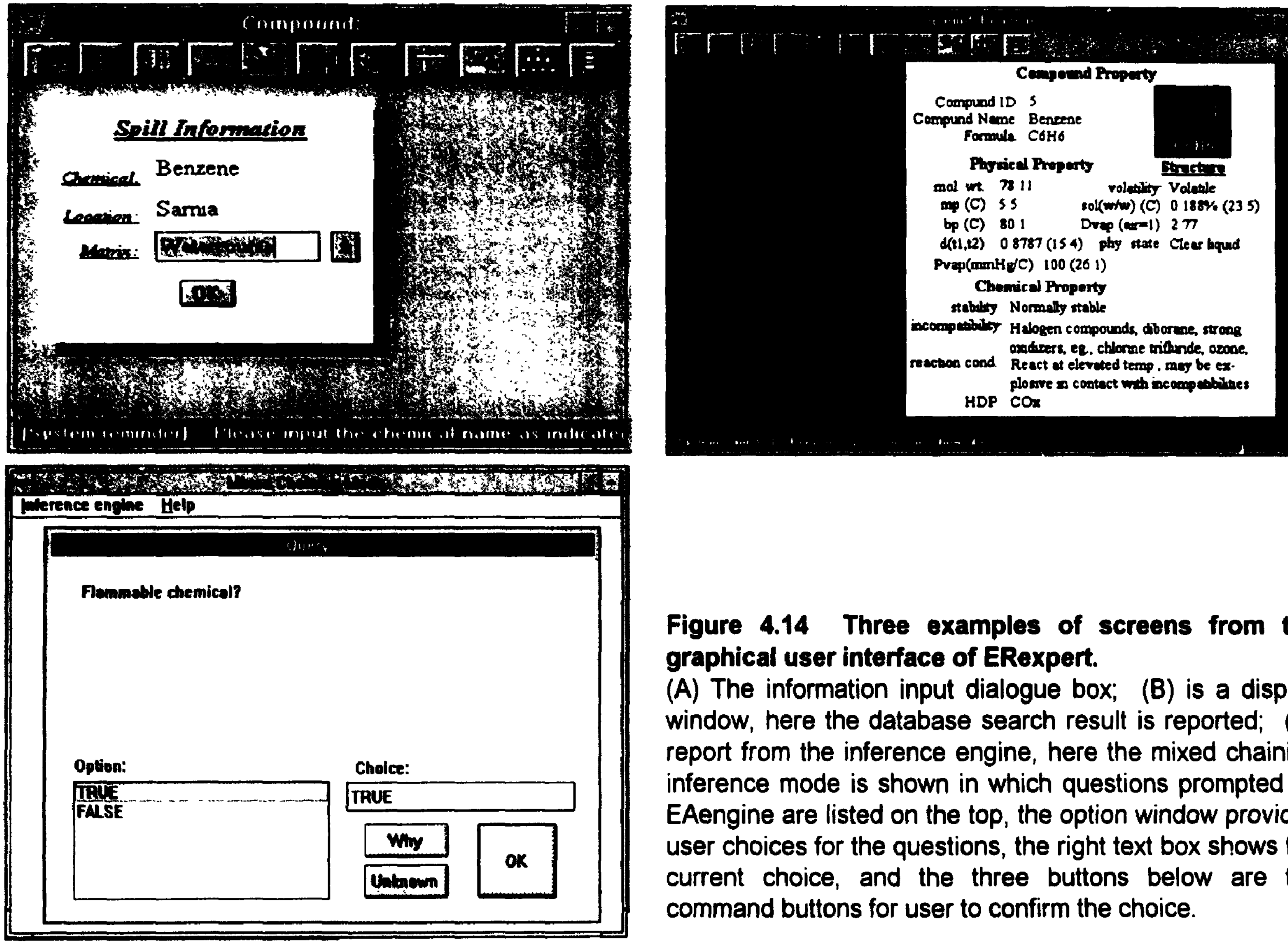


Figure 4.14 Three examples of screens from the graphical user interface of ERexpert.
(A) The information input dialogue box; (B) is a display window, here the database search result is reported; (C) report from the inference engine, here the mixed chaining inference mode is shown in which questions prompted by EAengine are listed on the top, the option window provides user choices for the questions, the right text box shows the current choice, and the three buttons below are the command buttons for user to confirm the choice.

answered, the particular conclusion that satisfies this question set is identified by the inference engine and displayed in a result dialogue box The forward chaining mode can be fast but is only appropriate if the user is certain that all the conditions can be identified correctly. In this process, the system will present an entire selection of symptoms to the user, from those selected the engine will conduct its reasoning based on the answers without further questioning. The backward chaining mode is the best choice if part of the knowledge matrix can be identified as the causes of a problem. This inference process is made possible because the arrangement of the knowledge in the KDM is goal oriented. Significantly fewer questions will be asked in this mode because only those questions that are related to the selected goal variable need to be answered. EAengine is capable of making suggestions even if incomplete information is used. In order to do this, the inference engine counts the number of hits (Q) in a partially fulfilled inferential process and compares this number to the number required to fire the rule (C). Then it calculates the probability of the suggested conclusion(s) and displays the result(s) as confidence factor [CF=(Q/C)x100%]. In the final display, the suggested answers are ranked according to the CFs. Once an inference process has been completed following the operation sequence shown in Figure 4.13, the user can either apply the suggestions to direct another search of the Factbase to gather more factual information or use them directly in the response planning.

Specific functions have been built into the **ERexpert** module and are described next. Access is provided to a directory of the regional response centers (RRCs). A notebook is available for recording information important to the response process, once clicked, a template will appear on the screen and the text written there can be systematically backed-up or saved-on-exit as files using *date* plus extension *txt* as filenames in a sub-directory in the **ERexpert** main directory (..\Diary). The multi-tasking ability inherent in MS Windows© enables this notebook to be left open and accessed during the whole response operation. The textual information written in the notebook can be readily integrated into the "Case-History" sub-database as a record of an individual case and an update function is embodied in the user interface for this purpose. These case records result from use of the **ERexpert** program and therefore will be evaluated against their application significance to decide whether to be included into the Case-History sub-database. This is a quick upgrade that only modifies the "Case-History" sub-database, the subject-oriented factual tables in the Factbase will not been changed. A full scale update of the Factbase requires the user to perform a series of modifications on each of the subject-oriented tables. This process requires not only the collection of factual information, but also knowledge of both use of MS ACCESS© and the structure of **ERexpert** database. Unlike the quick update of the "Case-History" sub-database, the full scale integration must be conducted by a knowledge engineer. A separate module (**DB**) is built and can be accessed through the **SPILLexpert** main window (Figure 4.12A).

4.8 APPLICATION EXAMPLES

This session assumes an accident involving TOLUENE occurred on unpaved ground near the side of a highway under normal weather condition. According to the operation sequence in Figure 4.13, **ERexpert** first searches the Factbase and because TOLUENE is not included in the current database, this search fails. If the inductive process is initiated and the mixed chaining mode is chosen, the user needs to answer a set of questions prompted by the system. The question sequence is as follows:

Knowledge Layer 1 (*CateID*):

Q1: Highly volatile chemical?

A1: NO

Q2: Liquid chemical?

A2: YES

Q3: Density (liquid/solid) less than water?

A3: YES

Q4: Flammable chemical?

A4: YES

Therefore, the inference process based on Knowledge Layer One reaches a conclusion and the system prompts for further reasoning through a dialogue box. As demonstrated in Figure 4.13, if further reasoning is required, the inference engine combines the current conclusion that the CateID is equal to Cate2B (category for flammable liquid chemicals), with new conditions in an effort to identify further conclusion(s).

Knowledge Layer 2 (*ContTech*):

Q5: Mechanical failure?

A5: NO (or previous mechanical failures have been fixed)

Q6: Spill happened on land?

A6: YES

Q7: Dry surface?

A7: Unknown (do not care since a liquid chemical was involved in the accident)

Q8: Cate2B chemical?

A8: YES (the previous inferential result)

Q9: Soft ground/river bottom?

A9: YES

Since TOLUENE belongs to a group of less volatile and flammable liquid chemicals, **ERexpert** suggests that it is possible that fire and explosion may occur following such an accident, therefore, if possible, proper coverage of the spill site is recommended to minimize the spreading of the vapor and no sparks should be allowed in the accident area. Next, possible containment and clean up procedures for the spill are given, these include a combination of the construction of earthen dikes and trenches, and the removal of bottom soil

Figure 4.15 shows the dialogue boxes in which the conclusion and explanation are given Detailed instructions about the usage of equipment and possible precautions may be found in the auxiliary tables in the Factbase accessed by the CateID.

It is important to note that the way in which the system proposes subsequent questions is based on a statistical analysis of the knowledge matrix, therefore, the order of the rules in the KBF file is a significant factor in deciding the overall performance of the inference engine. The general principle is to place the rules in the order of importance, the rules that are most frequently used should be put at the beginning of the KBF file (refer to Appendix 3). EAengine accepts three answer options: positive confirmation (YES, TRUE), negative confirmation (NO, FALSE), and no confirmation (UNKNOWN, DON'T KNOW). Sometimes, the last option can be interpreted as DON'T CARE. It is found that the way a question is answered may affect the number of questions asked by the inference engine though the final conclusion stays the same. This is because in order to completely represent the domain expertise, the same condition may occasionally be presented twice in the knowledge matrix. Therefore, in the inference process, especially when multi-step induction is involved, one question can appear more than once. Depending on the nature of the question, such questions can be answered differently. For example, the answer to question 7 in the example above can be either 'YES' (we had assumed normal weather condition for this session) or 'UNKNOWN' (don't care), but we find that the two different answers result in two totally different sequences of questions. A 'YES' to Q7 will generate a question sequence consisting up to 17 questions instead of 5 listed above when the answer was UNKNOWN in the second round of reasoning (Knowledge Layer 2), although the final conclusions are identical. We investigated why this happened by examining the chemical significance of the two different answers to the same question. In this example, a 'YES' to Q7 implies that the dryness of the ground becomes a significant factor

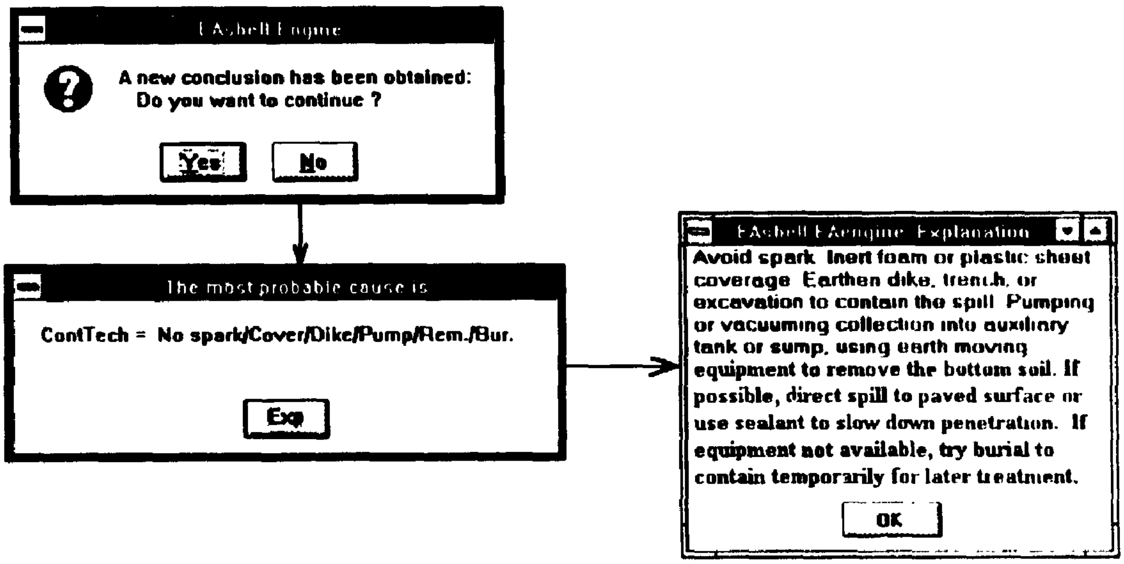


Figure 4.15 Dialogue boxes for the inference process used in ERexpert. The top dialogue box provides communication between the inference engine and the user and allows the user to decide whether a further search is required. The conclusion box shows a brief conclusion, and the right side dialogue box gives a more detailed explanation of the conclusion.

in considering potential response actions. Such a possibility forces EAengine to search deeper reasoning loops to consider the physical state of the spilled chemical(s) because solid chemicals may be fluidized by the wet ground and that may alter the selection of the response actions. On the other hand, wet ground may not affect the response dramatically for liquid chemicals. As a result, more questions are asked in order to get a clearer definition of the spilled chemical. In Table 4.8, more detailed results are listed from a thorough study of the relationship between the answer options and question sequences. The results also show that the order of the ASK in the user query section of a KBF file does not affect the question sequence chosen by the inference engine.

4.9 DISCUSSION AND CONCLUSION

4.9.1 Task characterizations

A number of expert systems have been developed in the chemistry domain [46-47]. These systems are concerned with three large fields of research: instrumental diagnosis, data interpretation and structure elucidation, and synthesis planning. Instrumental diagnosis deals mainly with verbal expression of human expertise, and the target domains are bound to specific instrumentation and are, therefore, clearly defined [25,33,35,48]. On the other hand, data interpretation and structure elucidation expert systems deal largely with numerical data in which a specific analytical instrument is involved and inter-discipline overlaps are not often observed [49-52]. Traditional planning-type expert systems in the chemistry domain have a similar character, for example, the determination of synthetic pathways for a given organic compound [53-55]. Recently, Olivero *et al.* [40] have described the development of an expert system to assist in the selection of experimental plans that ranks thirteen types of experimental design according to the suitability to the proposed project.

Table 4.8 Relationship between the answers chosen and the question sequences in a study session using the Toluene spill case.[e]

Sequence 1 [a]
(Random RULE order)

Q1: Mechanical failure?
A1: NO
Q2: Category 1A chemicals?
A2: NO
Q3: Sheltered area?
A3: NO
Q4: Category 1B chemicals?
A4: NO
Q5: Spill happened on land?
A5: YES
Q6: Dry ground surface?
A6: YES
Q7: Category 2A chemicals?
A7: NO
Q8: Category 3A chemicals?
A8: NO
Q9: Liquid state?
A9: YES
Q10: Flammable?
A10: YES
Q11: Volatile?
A11: YES
Q12: Category 4A chemicals?
A12: NO
Q13: Category 4B chemicals?
A13: NO
Q14: Category 4C chemicals?
A14: NO
Q15: Category 4D chemicals?
A15: NO
Q16: Category 4E chemicals?
A16: NO
Q17: Category 4F chemicals?
A17: NO
Q18: Category 3B chemicals?
A18: NO
Q19: Category 2B chemicals?
A19: YES
Q20: Soft ground surface?
A20: YES

Sequence 2 [b,1]
(Structured RULE order)

Q1: Mechanical failure?
A1: NO
Q2: Spill happened on land?
A2: YES
Q3: Dry ground surface?
A3: YES
Q4: Category 2A chemicals?
A4: NO
Q5: Category 3A chemicals?
A5: NO
Q6: Liquid state?
A6: YES
Q7: Flammable?
A7: YES
Q8: Volatile?
A8: YES
Q9: Category 4A chemicals?
A9: NO
Q10: Category 4B chemicals?
A10: NO
Q11: Category 4C chemicals?
A11: NO
Q12: Category 4D chemicals?
A12: NO
Q13: Category 4E chemicals?
A13: NO
Q14: Category 4F chemicals?
A14: NO
Q15: Category 3B chemicals?
A15: NO
Q16: Category 2B chemicals?
A16: YES
Q17: Soft ground surface?
A17: YES

Sequence 3 [b,c]
(Structured RULE order)

Q1: Mechanical failure?
A1: NO
Q2: Spill happened on land?
A2: YES
Q3: Dry ground surface?
A3: YES
Q4: Category 2A chemicals?
A4: NO
Q5: Category 3A chemicals?
A5: NO
Q6: Liquid state?
A6: UNKNOWN (don't care)
Q7: Category 2B chemicals?
A7: YES
Q8: Soft ground surface?
A8: YES

Sequence 4 [b,d,2]
(Structured RULE order)

Q1: Mechanical failure?
A1: NO
Q2: Spill happened on land?
A2: YES
Q3: Dry ground surface?
A3: UNKNOWN (don't care)
Q4: Category 2B chemicals?
A4: YES
Q5: Soft ground surface?
A5: YES

a Unstructured RULES in a random order in the KBF file. As a result, irrelevant questions are asked by the system.
b Structured RULES placed in the order of importance in the KBF file. This improves the system's performance by eliminating the irrelevant questions asked in Sequence 1.
c An answer of UNKNOWN (DON'T CARE) to any one of Q6 ~ Q8 of the questions asked in Sequence 2 will force the system to an abandon unnecessary 'deeper' reasoning loop to select the most obvious conclusion(s).
d This is the best answer sequence, and was illustrated in the text because it clears any ambiguity in the answers to force the system to identify the most obvious conclusion.
e Superscript 1 and 2: these two question sequences are used as examples in the text.

Nevertheless, the planning task by its nature can be a much more complicated process because it often involves factual information of both textual and numeric formats, as well as heuristic knowledge. While response to a spill is typically a planning task, it becomes more difficult because inter-disciplinary fields are involved and human experts in this area are specialized in only one of the many aspects and thus, are only proficient in one side of the problem. Consequently, for the **SPILLexpert** project, which aims at the development of an expert system capable of solving all planning tasks in response to spills, the development process becomes more difficult because the knowledge overlap magnifies an already complicated situation. One solution to this difficulty is to use segregated development and a systematic integration scheme, an approach first described by Kateman [56]. Using such a scheme, the general expert system framework may consist of many stand-alone modules, and the number is dependent on the application, in which each subsystem focuses on solving a sub-problem area of the problem domain. Van den Bogaert *et al.* [57] have also discussed the possibility of building an expert system incrementally by constructing separately developed subsystems into a general framework. The ESCA project has shown that such a process is possible. However, the need to fill the knowledge gap between different stand-alone expert systems may be difficult [58]. The development work on **ERexpert** has demonstrated that this scheme can be useful in reducing the complexity of a development target, especially when faced with a multi-disciplinary domain problem.

4.9.2 System update and knowledge base expansion

The knowledge base of the current **ERexpert** module employs 97 rules that are mainly used to deal with problems from three aspects: chemical categorization, containment methodology, and treatment-site selection. The menu-item 'HELP' is designed to allow users to request definitions of unfamiliar terms used by **ERexpert**. If **ERexpert** is to find its way into environmental applications, the program must be flexible and easy to integrate. It should be possible to add new knowledge or adapt the system according to changes in the application environment. The **ERexpert** prototype can be updated through separate refurbishing of the knowledge base without changing the master computer code. The knowledge base developed using the KDM encoding scheme is relatively easy to validate, and one may change the logical relations in certain parts of the knowledge base without upsetting the rest. An expert with little computer skill can modify the knowledge base by adding more knowledge or by rewriting an unclear part.

4.9.3 Limitations and perspectives

The **SPILLexpert** project focused on the knowledge acquisition process and the development of a knowledge base for an expert system prototype intended to assist the development of appropriate emergency responses to chemical spills. In this study, we have applied the compound categorization concept to characterize hazardous chemicals into four major categories and seventeen subgroups based on their environmental behavior. A subject-oriented Factbase containing the RHCs was developed using the Microsoft ACCESS® rational database system. With the help of causal analysis, heuristic knowledge related to chemical categorization was extracted and compiled into the Rulebase through the implementation of the KDM process. A fully functional knowledge base resulted that encompasses two parts, a Factbase that contains factual information, and a Rulebase used as a supplement to extend the coverage and applicability of the Factbase.

As an innovative knowledge encoding method, the KDM process has been described in detail in this chapter. A KDM process includes the following three steps: (i) causal analysis of the knowledge domain, (ii) transfer of the causal analysis results into the knowledge matrix and filling in the logical connections, and (iii) conversion of a filled KDM into production rules. One of the advantages of the KDM process is that it helps the knowledge engineer to visualize logical relations between conditions and conclusions that are otherwise difficult to express. The KDM scheme also provides flexibility for future modification of the logical relations between individual conditions and conclusions without upsetting the other connections in a knowledge table. The significance of this new knowledge encoding approach can be seen through this research, however, further application of this methodology will be necessary to verify the general applicability.

The **ERexpert** modular system is still very much in the research phase. The program currently covers most of the containment countermeasures and primary remedial actions in handling spill accidents. However, remedial procedures can be rather complicated and often involve dynamic processes that require mathematical modeling in order to simulate the processes of distribution, degradation, and detoxification of chemicals in the natural environmental. The current **ERexpert** model does not consider the effects of a complex environmental matrix or weather factors, and, therefore it is incapable of simulating the response processes dynamically. While mathematical modeling may be an effective approach in the prediction of the distribution of spilled chemicals, we expect to see major research advances to be made in the environmental research area so that the environmental behavior of chemicals, their disruptions to the eco-system, and their degradation processes in the environment, can be more quantitatively described. Kollig *et al.* [2] wrote that experimental conditions cannot be extrapolated to environmental conditions without considerable mathematical manipulation. Nevertheless, our study is valuable because **ERexpert** is not only a reference tool for field professionals dealing with chemical spills, but also provides the possibility for integration of domain expertise and relevant research progress into an expert system in which few applications have been reported. Although, the **ERexpert** modular system deals with only sub domains within the spill domain, we expect that the design protocol and knowledge encoding scheme developed through this study is generally applicable for subsequent development of other modular systems in the **SPILLexpert** project. In the next phase of project development, we plan to implement this protocol into the development of the ACmethod modular system. We anticipate that with the multi-tasking ability of MS Windows©, the difficulty in linking the different modular systems can be easily solved [35].

4.10 REFERENCES

1. Leonard, R. B. *Aviat. Spac. Environ. Med.*, **1993**, 64, 546-551.
2. Kollig, H. P.; Kitchens, B. E.; Hamrick, K. *J. Chem. Inf. Comput. Sci.*, **1993**, 33, 131-134.
3. Penmetsa, R. K.; Grenney, W. J. *J. Environ. Eng.*, **1993**, 119, 231-247.
4. Pierce, T. H.; Hohne, B. A. *Artificial Intelligence Application in Chemistry: ACS Symposium Series No. 306*; American Chemical Society: Washington, DC, 1986.
5. Hohne; B. A., Pierce, T. H. *Expert Systems Application in Chemistry: ACS Symposium Series No. 408*; American Chemical Society: Washington, DC, 1989.
6. Hushon, J. M. *Expert Systems for Environmental Applications: ACS Symposium No. 431*; American Chemical Society: Washington, DC, 1990.
7. Hushon, J. M. In *Expert Systems for Environmental Applications: ACS Symposium Series 431*; Hushon, J. M., Ed.; American Chemical Society: Washington, DC, 1990; Chapter 1, p 1.
8. Hushon, J. M. *Environ. Sci. Tech.*, **1987**, 21, 838-841.
9. Grinthal, W. *Chem. Eng.*, **1993**, 100, 179-182.
10. Olivero, R. A.; Bottrell, D. W. In *Expert Systems for Environmental Applications: ACS Symposium Series 431*; Hushon, J. M., Ed.; American Chemical Society: Washington, DC, 1990; Chapter 5, p 69.
11. Fang, H. Y.; Mikroudis, G. M.; Pamukcu, S. P. In *Expert Systems for Environmental Applications: ACS Symposium Series 431*; Hushon, J. M., Ed.; American Chemical Society: Washington, DC, 1990; Chapter 11, p 146.
12. Hirayama, H.; Wohlsen, R.; Brede, C. In *Expert System Applications in Chemistry: ACS Symposium 408*; Hohne, B. A.; Pierce, T. H., Eds.; American Chemical Society: Washington, DC, 1989, Chapter 16, p 200.
13. Schaum, J. L.; Segna, J. J.; Young, J. S.; Benes, C. M.; Muir, W. R. In *Expert Systems for Environmental Applications: ACS Symposium Series*

431; Hushon, J. M., Ed.; American Chemical Society: Washington, DC, 1990; Chapter 13, p 176.

14. Hanratty, P. J.; Joseph, B. *Comput. Chem. Eng.*, **1992,** 16, 849-860.
15. Subramanian, C.; Andreas, C.; Pandit, N. S. In *Proceedings of National Resources and Development Conference*; HMCRI, Anaheim, California, 1991.
16. Buelt, J. L.; Stottlemyre, J. A.; White, M. K. In *Proceedings of National Resources and Development Conference*; HMCRI, Anaheim, California, 1991.
17. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 945-955.
18. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 956-968.
19. Leonard, R. B. *Aviat. Spac. Environ. Med.*, **1993**, 64, 546-551.
20. Crittenden, G. *Hazadous Material Management*, **1995**, Aug./Sept., 14-18.
21. Breashears, D. D.; Whicker, F. W.; Hakonson, T. E. *Ecol. Appl.*, **1993**, 3, 590-593.
22. Levin, S. A. *Ecol. Appl.*, **1993**, 2, 103-106.
23. Breuker, J.; Wielinga, B. In *Knowledge Acquisition for Expert system*; Kidd, A. L. Ed.; Plenum Press: New York, 1987; Chapter 2, p 17.
24. Shank, R.; Riesback, C. R. *Inside Computer Understanding*. Lawrence Erlbaum Association: Hillsdale, NJ, 1981.
25. Zhou, T.; Isenhour, T. L.; Zamfirbleyberg M.; Marshall, J. C. *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 79-87.
26. Huang, G.; Stillman, M. J. An expert system shell under development in MJS' laboratory, unpublished results.
27. Settle, F. Jr.; Pleva, M. A. *Chemometr. Intell. Lab. Syst.*, **1991**, 11, 13-26.
28. Cancilla, D. A.; Huang, G.; Ma, S.; Stillman, M. J. GCMS-expert: An expert system for GC-MS data. In *Proceedings of the 40th ASMS Conference*, Washington, DC, June, 1992; p 993.
29. (a) Lahiri, S.; Stillman, M. J. AC-QC: A Quality Control Expert System and Its Interaction with AAcontrol and AAdiagnosis. In *Proceedings of the*

Technology Transfer Conference, Ontario Ministry of the Environment (MOE), Toronto, November 5-6, 1992; BP44. (b) Du, H., Lahiri, S.; Stillman, M. J. Diagnostic Expert Systems: An Icon-Based Expert System for Diagnosis of Problem GC data. ibid. BP44B. (c) Stillman, M. J.; Lahiri, S.; Huang, G. Design, Coding and Implementation of Expert System in Environmental Analytical Chemistry. ibid. BP44C. (d) Lahiri, S.; Stillman, M. J. Diagnostic Expert Systems: Encoding Chemical Knowledge in AAdiagnosis. In *Proceedings of the Technology Transfer Conference*, MOE, Toronto, November 25-26, 1991; p 685. (e) Zhu, Q.; Stillman, M. J. Design and implementation of ACselect. ibid. p 704. (f) Stillman, M. J.; Lahiri, S.; Zhu, Q. Design, Constraints and Implementation of Rules Within ACexpert. In *Proceedings of the Technology Transfer Conference*, MOE, Toronto, November 19-20, 1990; p 628. (g) Stillman, M. J.; Moussa, M.; Gasyna, Z. Development of ACexpert. 3. Rules in ACdiagnosis and ACmethods. In *Proceedings of the Technology Transfer Conference*, MOE, Toronto, November 20-21, 1989; p 254. (h) Stillman, M. J.; Cox, T.; Browett, W. R. Development of ACexpert. 2. Implementation of an Expert System for Automated Metal Analysis by AAS. In *Proceedings of the Technology Transfer Conference*, MOE, Toronto, November 28-29, 1988; p 195. (i) Browett, W. R.; Cox, T. A.; Stillman, M. J. Development of ACexpert. 1. Design of an Expert System for Automated Metal Analysis by AAS. In *Proceedings of the Technology Transfer Conference*, MOE, Toronto, November 30 - December 1, 1987; p 154.

30. Browett, W. R.; Stillman, M. J. *Prog. Anal. Spectrosc.*, **1989**, 12, 73-110.
31. Browett, W. R.; Cox, T. A.; Stillman, M. J. In *Expert System Applications in Chemistry: ACS Symposium Series No. 408;* Hohne, B. A.; Pierce, T. H., Eds.; American Chemical Society: Washington, DC, 1989; Chapter 17, p 210.
32. Stillman, M. J.; Huang, G.; Lahiri, S; Zhu, Q. ACexpert. Design and Implementation of ACselect, AAexpert and GC-MSexpert Systems that Aid

in the Analysis of Environmental Samples. In *Expert systems World Congress Proceedings*; Liebowitz, J., Ed.; Pergamon Press: New York, 1991; Vol. 4, 2645-2653.

33. Lahiri, S.; Stillman, M. J. *Anal. Chem.*, **1992**, 64, 283A-291A.
34. Du, H.; Stillman, M. J. *Anal. Chim. Acta*, **1994,** 296, 33-41.
35. Du, H.; Lahiri, S.; Huang, G.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 21-31.
36. Zhu, Q.; Du, H.; Stillman, M. J.; Plomley, J. B.; March, R. E. GCMSdiagnosis - An expert system for diagnosis of problem GC-MS data. *Anal. Chim. Acta*, **1995**, submitted.
37. Lahiri S.; Yuan B ; Stillman, M. J. *Anal. Chem.*, **1994**, 66, 2954-2963.
38. Lahiri S. *Ph.D. Thesis*, The University of Western Ontario, London, Ontario, 1994.
39. Zhu, Q.; Stillman, M. J. ACexpert: Design of a User Interface. In *Proceedings of the 74th Canadian Chemical Conference and Exhibition,* Hamilton, Ontario, June 2-6, 1991; 61 AN-E1.
40. Olivero, R. A.; Seshadri, S.; Deming, S. N. *Anal. Chim. Acta*, **1993**, 277, 441-453.
41. Penninckx, W.; Smeyers-Verbeke, J.; Massart, D. L. *Analy. Chem. Acta*, **1993**, 282, 417-422.
42. Zhu, Q.; Stillman, M. J. Knowledge development and system design for SIRS: An expert system for use in response to emergency chemical spill. In *Proceedings of the Technology Transfer Conference*, Ontario Ministry of the Environment (MOE), Toronto, Ontario, November 5-6, **1992**; BP44A.
43. Dawson, G. W.; Shuckrow, A. J.; Mercer. B. W. In *Hazardous Chemical Spill Cleanup: Pollution Technology Review No. 59*; Robinson, J. S., Ed.; Noyes Data Corporation: New Jersey, 1979; Chapter 2, p 23.
44. Edwards, M.; Cooley, R. *CABIO*, **1993,** 9, 657-665.

45. Kidd, A. L. In *Knowledge Acquisition for Expert system: A Practical Handbook*; Kidd, A. L., Ed.; Plenum Press: New York, 1987; Chapter 1, p 1.

46. Zhu, Q.; Stillman, M. J. Expert Systems and Analytical Chemistry: Recent Progress in the ACexpert Project. *J. Chem. Inf. Comput. Sci.*, **1995**, in press.

47. Jakus, V. *Collect. Czech. Chem. Commun.*, **1992**, 57, 2413-2451.

48. Settle, F. A., Jr.; Walter, P. J.; Kingston, H. M.; Pleva, M. A.; Snider, T.; Boute, W. *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 349-353.

49. Pyo, D. *Vibr. Spectrosc.*, **1993**, 5, 263-273.

50. Zupan, J.; Buydens, L.; Peeters, A.; Massart, D. L. *Chemometr. Intell. Lab. Syst.*, **1990**, 8, 43-51.

51. Warr, W. A. *Anal. Chem.*, **1993**, 65, 1045A-1050A.

52. Warr, W. A. *Anal. Chem.*, **1993**, 65, 1087A-1095A.

53. Hendrickson, J. B.; Parks, C. A. *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 209-215.

54. Dogane, I.; Takabatake, T.; Bersohn, M. *J. Royal Netherlands Chem. Soc.*, **1992**, 111, 291-296.

55. Pollet, P. *New J. Chem.*, **1990**, 14, 957-963.

56. Kateman, G. *Trends Anal. Chem.*, **1984**, 3, p VII.

57. Van den Bogaert, B.; Morsink, J. B. W.; Smit, H. C. *Anal. Chim. Acta*, **1992**, 270, 107-113.

58. Buydens, L.; Schoenmakers, P.; Maris, F.; Hindriks, H. *Anal. Chim. Acta*, **1993**, 272, 41-51.

APPLICATION CASE II: ANALYTICAL CHEMISTRY

CHAPTER 5
DEVELOPMENT OF A DIAGNOSTIC EXPERT SYSTEM FOR GC-MS

5.0 INTRODUCTION

In recent years, the progress made in quadrupole ion trap mass spectrometry has been quite extraordinary [1]. In some favorable cases, the detectable mass range exceeds 70,000 Da [2], a mass resolution range from 10^6 ~ 10^7 [3-4], and detection limits down to the attomole (10^{-18} mol) level [1], have been reported. Various commercial ion trap mass spectrometers (ITMS) are available, and with the help of computer software, additional features, such as ion isolation, alternative scan functions, resonance excitation with a supplementary radio frequency, and tandem mass spectrometry (MS^n, $n \geq 2$), are made possible [5]. In environmental analysis, the use of gas chromatography with a mass spectrometer detector (GC-MSD) is preferable in many situations because the MSD can greatly enhance the method specificity and lower the detection limit. For example, the ITMS instrument has been used in the determination of volatile organic compounds in drinking water [6], in the detection of trace level herbicides in surface and ground waters [7], and in the characterization of chemical waste sites [8].

Mass spectrometry based on the quadrupole ion trap, especially the tandem quadrupole mass spectrometer, is a relatively new technology In the past several years, many research studies have been undertaken to further the study on the chemical, physical and electromagnetic properties of the trap, the behavior of ions in the trap, as well as to investigate new operational modes that could incorporate such newly arrived knowledge into applications in the field of mass spectrometry. For instance, Williams *et al.* investigated the effects of scan direction on mass resolution in resonance ejection ion-trap mass spectrometry [9] Guan and Marshall studied equilibrium space-charge distribution in the quadrupole ion trap [10]. A study on non-linear resonance effects during ion storage in a quadrupole ion-trap has been reported [11]. And a series of investigations on a new operating mode of quadrupole ion-trap mass spectrometry have been conducted by Zerega, *et al.* [12-14]. ITMS has been used to study the analysis sensitivity of benzophenone and resulted in a lower limit of detection [15], and for ambient air analysis [16]. More recently, applications of GC tandem mass spectrometry using a quadrupole ion-trap detector has been employed in the detection of trace amount of tetrachlorodibenzo-*p*-dioxins (TCDDs) and N-nitrosodimethylamine (NDMA) in complex environmental matrices [5,17]. At the present stage, the use of ITMS as a single mass-selective detector for gas chromatography is well established, however, extension to the tandem mode is not straightforward [18]. Among other problems, a practical difficulty is to design a composite broad band isolation waveform (BBISO) with multi-fragment of resonant frequencies that will be based specifically on the analyte ions (the selected parent ions). While reduction of most of the matrix background by resonance ejection is possible, construction of a suitable BBISO requires extensive foreknowledge and operating experience since the space-charge problem could affect the selection of resonance frequencies even if a desired m/z ratio has been determined. Yet, such knowledge may not be readily available for application as the whole field of ion-trap mass spectrometry is under research and constant development.

Computer-based expert systems, through the integration of the knowledge of a rather narrow field of study, are able to provide solutions for domain specific problems. The use of computer hardware and programs has been common in mass spectrometry [19]. In fact, numerous mass spectrometric methods and applications would be impracticable without progress in micro-electronic and computer science [1,5]. Knowledge-based expert systems are amenable to

ITMS because this technique is complex, consisting of multiple instrument modules that require constant attention and knowledge in order to operate the instrumentation efficiently.

In this chapter, we describe the development work on an expert system prototype to help in diagnosing problem GC-ITMS data. The diagnostic expert system is built for use on a quadrupole ion storage mass spectrometer (QISMS™, trademark of Varian Associates, Inc.). We believe that fast and reliable detection and diagnosis of faults in analytical instruments, such as the case addressed in this chapter, requires a high degree of knowledge and experience. Therefore, automating the diagnosis will make the knowledge more accessible and allow verification of the coherence of expertise and reasoning of human expert(s) [20].

5.1 DESCRIPTION OF THE KNOWLEDGE DOMAIN

5.1.1 Definition of the GC-MS analysis procedure

During analysis by ion-trap GC-MS, several steps can be distinguished and are described as follows:

(1) An initial method (i.e. flow rate, column temperature, ionization mode and time, scan function, resonance frequency, etc.) is selected based on the information of the sample to be analyzed.

(2) Next, a method optimization step is required to refine the so-called 'first guess' method. Criteria are primarily given to the method specificity in order to achieve the required separation (chromatographic behavior), and the sensitivity (detecting behavior). In this study, it is the behavior of the ion-trap mass spectrometric detector.

(3) Finally, a method validation step is necessary to characterize the precision, accuracy, and the limitation of the chosen method. The performance check is focused on reproducibility of the analytical results.

A schematic view of the steps involved in a GC-MS analysis is given in Figure 5.1. For any instrumental analysis, optimization of the operating parameters is an important step in a method development process. Whenever a problem arises, the operator (not always a skilled analytical chemist as often the case in industrial laboratories) will have to be able to detect the fault(s), diagnose the possible cause(s), and ultimately correct the wrong setting(s). Such operation optimization/fault diagnosis is one of the most difficult tasks faced in reality, which demands a fairly thorough knowledge of the theoretical background of the instrumentation, as well as considerable amount of hands-on experience. In practice, often an expert is consulted for solution(s). However, the availability of such required expertise is always questionable. On the other hand, knowledge-based computer programs have demonstrated efficiency in processing and representing verbally formatted knowledge [21]. Because of the accuracy, consistency, reliability and availability, expert system technology offers a promising alternative to address this type of complicated problem, in which considerable amount of preknowledge and experience is involved [22].

5.1.2 GC-ITMSD system and signals generated

In order to perform fault detection and diagnosis on a gas chromatography with an ion-trap mass spectrometric detector (GC-ITMSD), some specific features of the instrument and related theories have to be considered. Figure 5.2 is a schematic diagram of the Varian Saturn™ GC-MS system (QISMS™) for which the diagnostic expert system prototype has been developed [23]. The Saturn™ system contains three major components: a gas chromatograph, a mass spectrometric detector, and a data handling system.

The principle of GC-MS analysis lies in the separation of target analytes from a complex matrix, both chromatographically and mass spectrometrically, and in the subsequent measurements of the intensities (quantitative) and the interpretation of patterns (qualitative/structural) of the fragmented ions (m/z). A

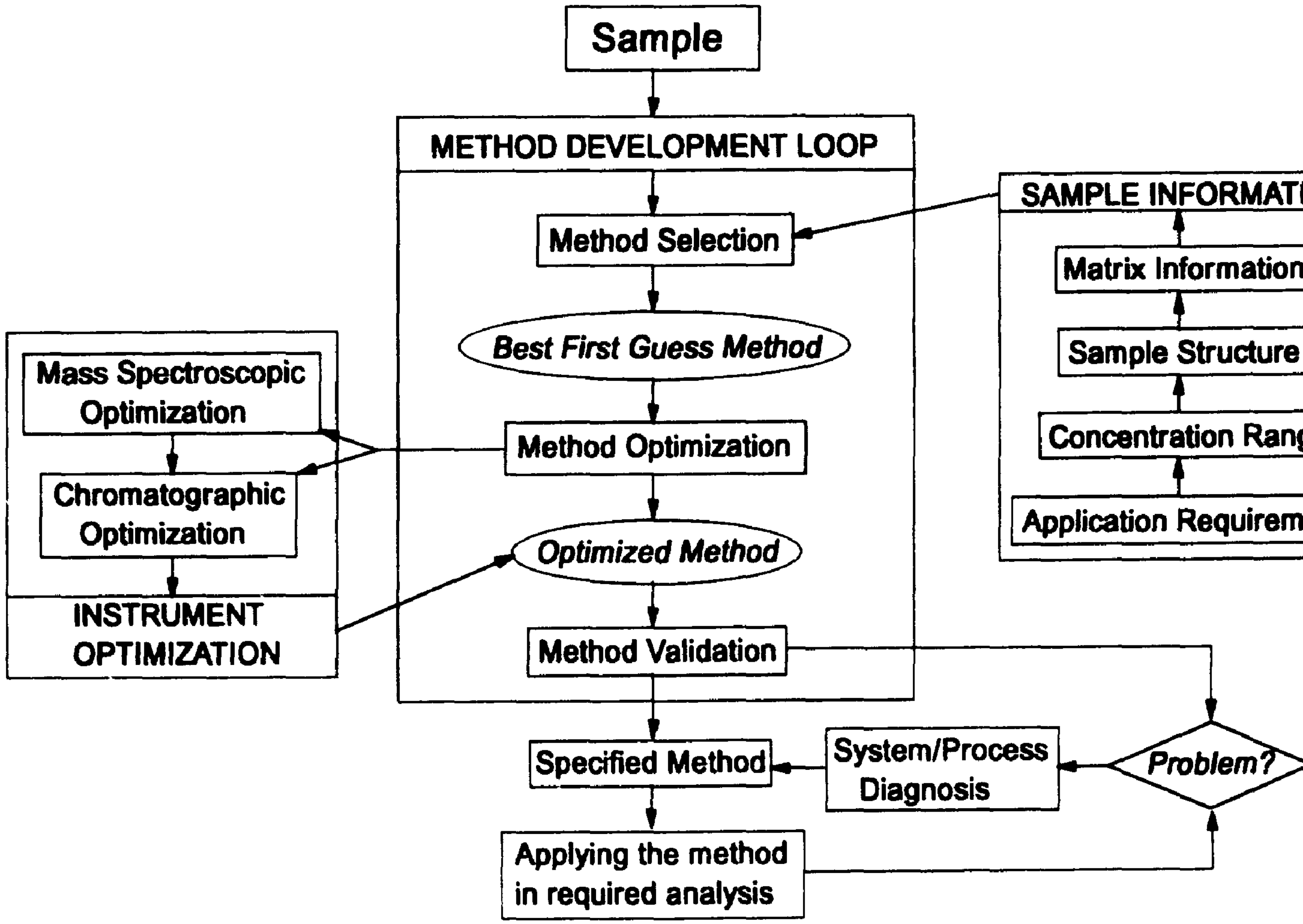


Figure 5.1 Schematic Illustration of the GC-MS Analytical Process.
System and process diagnosis is an important step in a process of GC-MS analysis Any unexpected problems need to be solved in this step before application of the chosen method in real analysis.

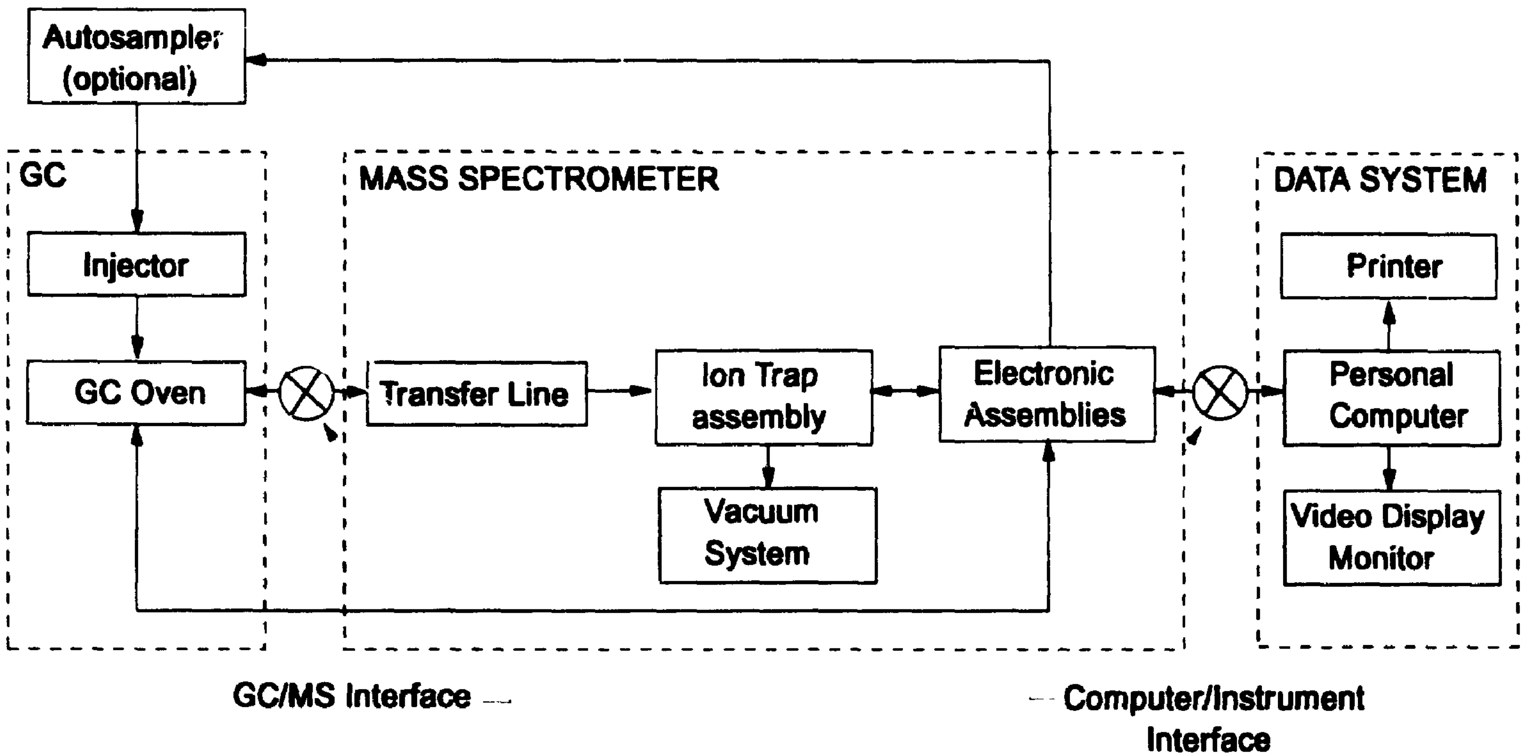


Figure 5.2 Schematic Diagram of the SATURN™ GC-MS System
The SATURN™ system contains three major components: a gas chromatography, a mass spectrometric detector, and a data handling system. The diagnostic expert system, GCMSdiagnosis, has been developed based on this instrumentation.

GC-ITMSD system can deliver three basic types of analytical signal:

(1) A total ion chromatogram (TIC) that provides both qualitative and quantitative information of the target sample. The quality of data is determined by the matrix complexity as well as the effectiveness of a chosen chromatographic separation procedure. A total ion chromatogram obtained from a GC-ITMSD system is similar to a gas chromatogram acquired on a GC-FID system.

(2) A single ion chromatogram that is the result of a mass selective operation. Based on a user-defined mass-to-charge ratio (m/z), qualitative and quantitative information about specific component(s) in a sample matrix can be obtained.

(3) A mass spectrum provides fragment pattern of sample ions and therefore, reveals certain degrees of structural information about an analytical target.

A tandem mass spectrometer (MS^n, $n \geq 2$) repeats the mass spectrometric process n times, and therefore is capable of providing fragment patterns for the offspring ions (daughter ions) based on a selected ion (parent ion). This process provides unprecedented structural information by eliminating most of the background ions so that the method specificity and sensitivity is sharply enhanced.

5.1.3 Brief theory about ion-trap mass spectrometry

Our group has previously developed a diagnostic expert system for detection of faults in gas chromatographic data, and the relationship between instrument parameters and the observed symptoms have been investigated fully [24-25]. In this chapter, we focus our attention on the mass spectrometer part of the instrument and a brief theory that is considered to be fundamental in understanding the ion-trap is given in this section.

The quadrupole ion-trap device consists of a ring electrode located between two hyperboloid end-cap electrodes. For an ion to be trapped, its trajectory must be stable both radially and axially. Two dimensionless quantities, a_z and q_z, are used to define the working points of an ion inside an ion-trap. If the working points fall within the stability diagram (a pseudo-potential well) and providing that the kinetic energy of the ion does not exceed the trapping potential, the ion is then trapped. When an ion-trap is operated in Mode II (see Appendix 4), the relationship between an ion (m/z), its working point in the trap (a_z, q_z), and the potential applied (V, U) is given as follows:

$$a_z = -2a_r = \frac{-8eU}{mr_0^2\Omega^2} \qquad (5.1)$$

$$q_z = -2q_r = \frac{-4eV}{mr_0^2\Omega^2} \qquad (5.2)$$

where, V is the zero-to-peak amplitude of the applied potential Φ_0, U is the d.c. component of the applied potential Φ_0, and Ω is the frequency of the *r.f.* voltage Refer to Appendix 4 for a brief summary of the principle theory about the ion trap mass spectrometry.

According to the above equations, the gaseous ions stored in an ion-trap can be destabilized in the same order as their mass-to-charge (m/z) ratios by scanning the amplitude of the *r.f.* potential applied to the ring electrode, continued with a supplementary *r.f.* voltage of amplitude (6 - 10 volt) $V_{(0\text{-}p)}$ with a fixed frequency (Ω) applied to the end-cap electrodes. When the *r.f.* drive potential is ramped to come into resonance with the axial secular frequencies, the ions will be excited rapidly, and will be ejected from the trap when their kinetic energies exceed the trapping potential well-depth. For example, let us consider three ion species of similar m/z ratios, (A-2), A, and (A+2). Assuming the initial working points of all three ion species lie inside the stability diagram on the q_z axis, and the ion A needs to be selected. By changing the voltage of the *r.f.* component (V), the working points of three ion species can be moved along the q_z axis until the working point for the ion (m/z = A) lies directly below the upper apex of the stability diagram. Next, a d.c. component with an appropriate value (U) is applied to locate this ion just inside the upper apex, while the working points of the rest ions (m/z = A$\pm$2) exceed the boundaries of the stability diagram from both sides, and are therefore, lost axially and radially, respectively. So that the stable trajectories are maintained only for those ion species whose working points lie within the apex, all others are ejected Finally, the *r.f.* and d.c. voltages are then reversed to move away from the boundaries and restore the working points of the isolated ion species to somewhere near the center of the stability diagram. When an ion-trap is operated in the tandem mode, this whole process can be repeated and the secondary ion species are therefore isolated in a similar manner. Unlike most of the tandem mass spectrometers which are "tandem in space", a tandem mass spectrometer using an ion-trap device is "tandem in time" because the ion-tap can be operated in a pulse mode by employing a series of discrete steps. A detailed description about the theory and practice of ion-trap mass spectrometry has been given by March [1,26].

5.2 EXPERIMENTAL

5.2.1 Instrumentation

Our laboratory is equipped with a Varian SATURN™ II QISMS™ system with a Varian GC3400, and a Varian Model 8100 autosampler An initial prototypic diagnostic system was built based on this instrument. The upgrade of this diagnostic system amenable to the GC-MSMS process was achieved through collaborations with Dr. March's group at Trent University Several field trips were made to the laboratories at Trent University, and at the MOEE (Ontario Ministry of the Environment and Energy), both facilities are furnished with a Varian SATURN™ III QISMS™ system, each equipped with a waveform generator, a Varian 3400 GC, and a Varian Model 8200 autosampler SATURN™ Revision C software was used for data acquisition while a prototype software package (QISMS™ version 1.0) was employed for the construction of the custom scan functions required for target analyte ion isolation and subsequent collisionally activated dissociation (CAD) of the isolated ion species. All the experimental data used to modify the diagnostic expert system prototype was acquired on these two Varian SATURN™ III GC/QISMS™ systems.

5.2.2 Computer and software

All the computer work was carried out on a 20 MHz Intel 80386-based computer, with 8 megabytes of RAM, two 150-megabyte hard disks, a 1.44-megabyte floppy drive, and a super VGA graphic monitor. The diagnostic program was developed for use in the Microsoft Windows© 3.1 operating environment. Tools employed in the development of this project were all Microsoft Windows© based applications, include: Microsoft Visual Basic© (ver 3.0, professional edition), an object-oriented program language that was used to write a graphical, self-explanatory user interface; and EAshell©, an expert system shell that is fully compatible with Microsoft Windows© [27].

EAshell© provides the inference engine (EAengine) for the program that is compiled as a dynamic linked library (dll). Inference processes can be initiated by function calls made to the engine. Three inference strategies are supported by EAengine, namely forward, backward, and mixed chaining. The system user is provided with options to select different knowledge bases and inference strategies. Once activated, EAengine uses the current knowledge base (rules) to try and solve a user specified problem. Sometimes additional questions may be asked by the inference engine in order to clarify ambiguities that occurred during an inference process.

5.3 RESULTS AND DISCUSSION

5.3.1 Components in the GCMSdiagnosis expert system

The primary objective of this project was to develop an expert system that can assist operators in fault detection and diagnosis in their routine operation of the GC-QISMS™ instrumentation. A general shell, named GCMSdiagnosis comprises four sub-systems that can be used for data handling and subsequent diagnostic uses. The structure of the system is illustrated in Figure 5.3. The data handling module, SPECview, is used to manipulate GC-MS analytical signals. SPECview is able to access directly both chromatographic and mass spectrometric data files saved in the Varian data format. Two major functions are built into this sub-system, one is to open/view data numerically, the other is to display both gas chromatograms and mass spectra. GCdiagnosis is a previously developed expert system for diagnosis of problem data acquired through a gas chromatographic analysis, and is incorporated into the GCMSdiagnosis shell as a sub-system. QISMSdiagnosis, stands for QISMS™ diagnostic expert system, is the diagnostic module for fault detection for the quadrupole ion-trap mass detector part of the instrumentation. The module DIAGplatform uses a quantitative variable to provide the QISMSdiagnosis program with a trouble-shooting starting point. We chose to use conversion efficiency as the guideline for the diagnostic process. In a tandem mass spectrometric process, the conversion efficiency (CE%) is defined as a fragment ion signal intensity expressed as a percentage of initial parent ion signal intensity. Threshold CE% values on standards and samples can be entered into this module. Whenever a CE% drops below the corresponding threshold value during a GC-MSMS analysis, DIAGplatform reports the problem, and subsequent diagnosis can be initiated to detect the cause(s) of the problem.

Each of the subsystems has been developed individually. The modules are connected together through a graphical user interface. Because these modules have been developed using a subject-oriented structure, it is easy to integrate

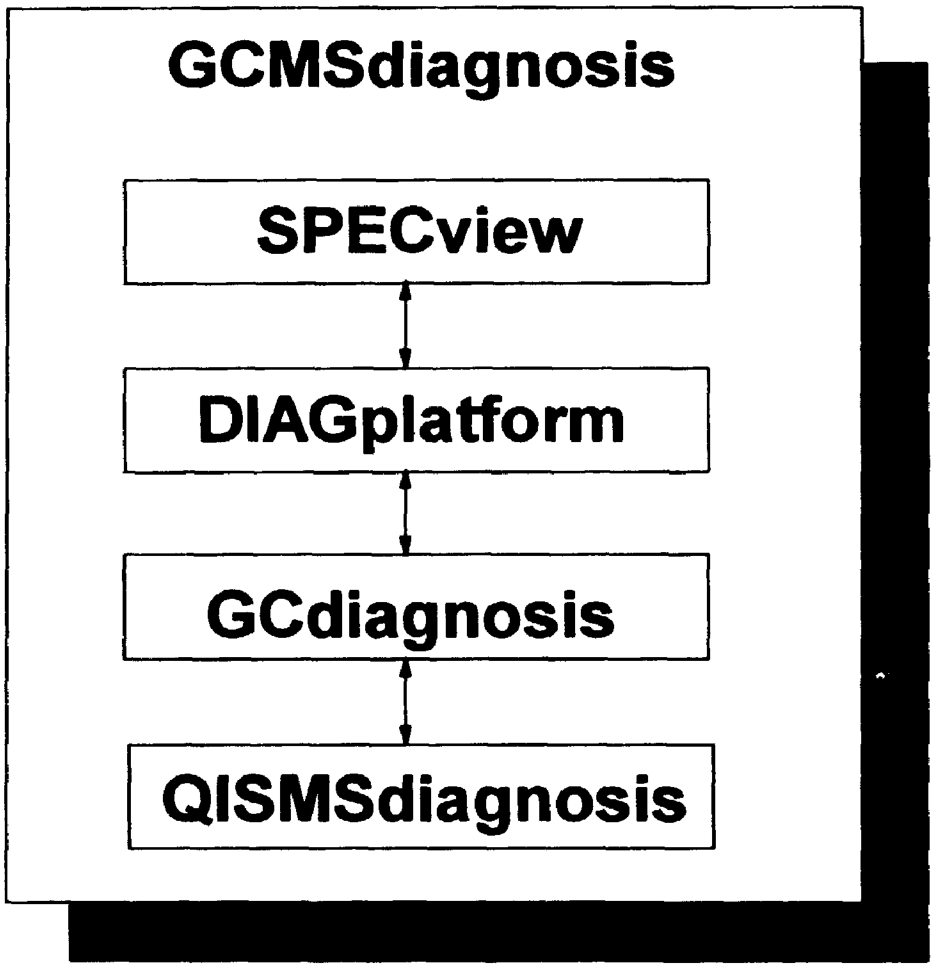


Figure 5.3 The structure of the GCMSdiagnosis program. GCMSdiagnosis consists of four sub-systems that can be used for data handling and subsequent diagnostic purposes. These modular expert systems can be used independently for specific tasks, or collectively as a complete package for GCMS data handling and problem diagnosis.

them into a common shell (GCMSdiagnosis). The order of operating the GCMSdiagnosis shell is flexible and can be decided by the system user or based upon the priority of a problem. For example, when the cause(s) of a problem can be isolated and related to a particular part of the GC-MS instrumentation, a single module can be used to solve the problem. More often, however, when no specific cause(s) can be identified, a complete diagnosis is required that runs each module consecutively until the cause(s) is detected or halted by the system user. GCMSdiagnosis is capable of giving suggestions even if incomplete information is used. This means that during a diagnostic process, if a user can not supply the system with all the symptoms (observations) required for a positive fault detection, GCMSdiagnosis is able to 'guess' the most probable cause(s) based on the symptoms provided by the user. In order to do this, EAengine counts the number of hits (Q) in a partially fulfilled inferential process and compares this number to the number required to satisfy a rule (C). It then calculates the confidence factors [CF=(Q/C)x100%] of the suggestions and displays the results along with their probabilities according to the rank.

5.3.2 Results from knowledge acquisition

Figure 5.4 summarizes the steps involved in the use of a KDM to process the heuristic knowledge of a defined problem domain. We have employed this process in several previous development studies [24-25,28-31]. Taking this route, we carefully studied the theoretical background involved in a GC-ITMSD analysis (refer Section 5.2 and Appendix 4), and collected all the possible faults known to us that are either observed in our routine analyses, recorded through discussions with other field experts, or extracted from text books and operational manuals. As a result, we obtained a primary diagnostic knowledge set. It should be noted that this knowledge set was not well defined and has various overlapping aspects, and even contradictions. Also different human experts have different habits in thinking about and solving a problem. This inevitably is

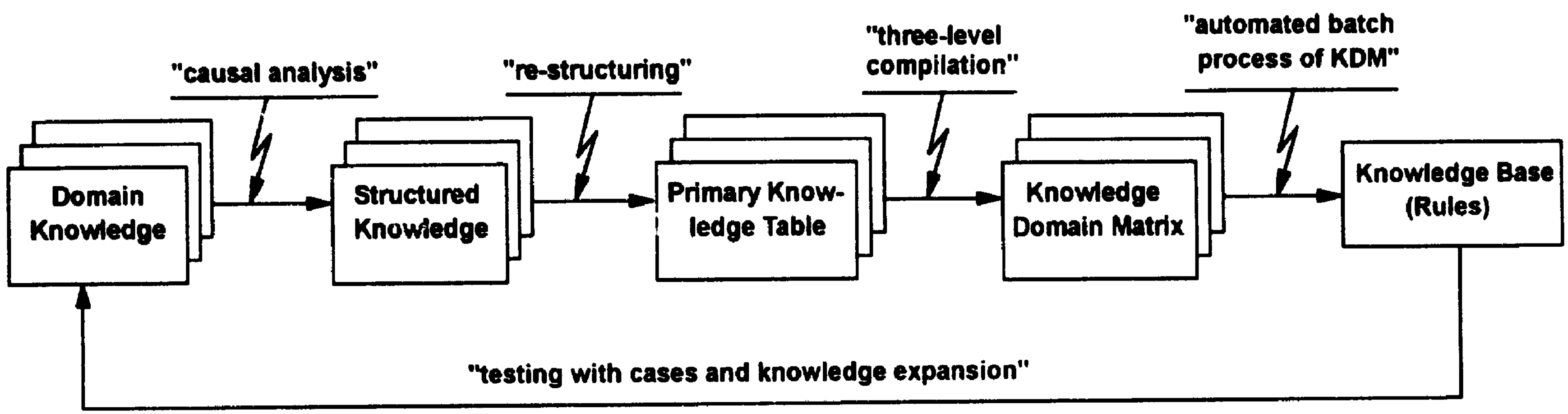


Figure 5.4 The steps involved in the use of the Knowledge Domain Matrix (KDM) to process heuristics in a defined problem domain.

A KDM scheme takes four steps: (1) Through causal analysis, knowledge of the problem domain is structured and causes and results are clearly identified; (2) The structured knowledge is translated into a tabular format, and forms the primary knowledge table; (3) The primary knowledge table is compiled through three cycles and logical connections are created; and (4) The knowledge represented in a KDM is reformatted into production rules through an automated batch process.

reflected in the recorded knowledge set. Therefore, a systematic approach is needed to analyze and modify the primary knowledge set. The structured knowledge set acquired through causal analysis is listed in Table 5.1. 30 symptoms and 37 possible causes that are frequently encountered in routine operations of the Varian SATURN™ GC/QISMS™ instrument have been identified. This structured knowledge is used as the training set to develop the GCMSdiagnosis program.

Next, the training set is entered into a blank matrix and three rounds of knowledge compilation are performed on the target matrix. The causes in a training set can be characterized into different sub-groups according to their functions in the analytical process. For diagnosis of GC/QISMS™, the causes can be divided into seven sub-groups: sample preparation, system connection, injector and injection, column, carrier gas, ion trap device, and data processing and communication. As illustrated in Figure 5.5, the above described operations have resulted in a domain knowledge matrix (KDM) with logical connections filled in corresponding cells. Notice that 7 more causes are listed as conclusions in the KDM. This is because one problem or fault may give different symptom combinations. Such cases are defined as parallel cases, and 4 more conditions (as indicated in Table 5.1 by parentheses) are added to the KDM in order to distinguish the resulting parallel cases. Low case letters in the KDM indicate an 'OR' connection, a logical value implies that either A or B plus a number of additional symptoms can result in a given situation. In other words, either one of A and B, not both, forms part of the premise necessary to reach a conclusion.

Computers do not work with physical objects, they invariably work with symbolic formats. In computer-based problem-solving process, an efficient representation and use of domain knowledge by the program is a very important aspect. The knowledge represented in a KDM needs to be translated into specific logical expression before EAengine can be used in the inductive processes. EAengine is designed to understand logical expressions using a rule

Table 5.1 List of frequently encountered symptoms and causes in routine analysis by GC-QISMS™.

Symptoms	Causes
1. poor resolution	1. contaminated GC eluting
2. low sensitivity	2. system leaking
3. unresolved peaks	3. severe system leaking
4. band broadening	4. severe leaking at col./septum
5. peak size change	5. leaking at col./septum
6. leading peaks	6. inject port temp. not hot enough
7. peak tailing	7. split ratio too high
8. t_R changing from run to run	8. injection speed too high
9. baseline drifting	9. syringe defect./partially plugged
10. high boiler severe tailing	10. septum bleeding/coring
11. high background noise	11. improving injection using SPI
12. severe water peaks (h_{18}/h_{19}<10:1)	12. column bleeding
13. air peaks (m/z: 28, 32, 18)	13. column damaged, replace
14. silicone peaks (m/z: 207, 73)	14. column degraded, replace
15. noise decrease as col. temp. decrease.	15. col. temp./program not optimized
16. ion clusters in whole mass range	16. column overload
17 short ionization T (<15,000 us)	17. mass transfer line temp. low
18. increasing GC temp. reduce problem	18. trap contaminated, condition overnight
19. RF tuning difficult	19. trap contaminated severe, cleaning
20. (M+1) peak significant	20. ion trap temp. low
21. abnormal AGC target value	21. manifold temp. low
22. abnormal electron multiplier emission current	22. AGC target value high
23. conditioning instrument component effective	23. wrong reagent gas
24. ionization time varies	24. ion neutralization reaction
25. weak daughter ion intensity	25. poor scan function
26 parent ion after tandem MS	26. wrong AGC target value
27. no signal	27. filament current small
28. can't communicate with ITMSD	28. elec. multiplier GAIN drop off
29. can't communicate with GC	29. incorrect AM amplitude
30 split injection mode	30. carrier gas flow high
31 (chemical ionization mode)	31. unstable carrier gas flow
32. (acidic or basic polar sample)	32. poor carrier gas quality
33. (ultra-trace analysis)	33. dirty plumbing
34 (instrumental malfunction)	34. active surface in inj./col.
	35. matrix effects
	36. below detection limit
	37. communication failure, detach/locate

SYMPTOMS / CAUSES	Leaking					Injector/Injection								
	1.Increasing GC temp.(250 C), IT less than 5000 indicating contaminated GC eluting. To locate a single source, run again be more specific.	2.System leaking.	3.System leaking (severe).	4.Carrier gas leaking at column/septum (severe).	5. Carrier gas leaking at column/septum.	6.Inject temp. not hot enough to vaporaze	7.Split ration too high.	8.Injection speed too high, adjust 1.0 ul/sec.	9.Syringe defective:partially plugged.	10.Septum bleeding coming with typical discrete ion clusters.	11.Improving injection tech using SPI.	12.Improving injection tech using SPI.	13 Column bleeding	14.Column damaged, replace.
1. high background noise masking sensitivity	T	T			T					T			T	T
2. noise level decrease as col. temp. decrease	t									t			t	
3. ion clusters throught the whole mass range										F			T	
4. silicone peaks (m/z: 73, 207, 281, 327 int. decs.)										T			[illegible]	
5 severe water peak (H18/H19 << 10:1)		T	T											
6 air peak (m/z: 28, 32, 18 broadened)		T	T											
7. conditioning suspected inst. component effective														
8. short ionization time (<15,000 us)	T													
9. RF tuning difficult														
10. low sensitivity			T	T	T	T	T		T		T			
11. acidic or basic polar sample								T			T	T		
12. weak daughter ion intensity														
13. chemical ionization mode														
14. (M+1) peak significant														
15. normal AGC target value (5,000 to 20,000)														
16. regular elec. multiplier emis. current (20 to 50 uA)														
17. unresolved peaks														
18 retention time changing from run to run			T	T			F		F					T
19 high boiler severe tailing						T								
20. band broadening						t		T						T
21. split injection mode							T							
22. increase GC zone temp. reduce the problem	F					T								
23. peak size change from run to run				T					T					
24. peak tailing								t						
25 poor resolution								T				T		
26. ionization time varies														
27 leading peaks														
28. baseline drifting		T			T									
29. no signal														
30 instrumental mulfunction suspected														
31 trace amount of sample (10^2 fg level)														
32. Left over parent ion after tandem MS process														
33 can't communicate with ITM/SD														
34. can't communicate with GC														

Column			Ion Trap															Carrier		
15.Column degradated, replace.	16.Column temp./program not optimized.	17.Column overload(>several hundred ng).	18.Mass transfer line temp. too low.	19.Ion-trap containm., conditioning overnight.	20.Ion-trap contaim.(sever), cleaning ion trap.	21.Ion-trap temp. too low, adjust 200-250 C.	22.Increase manifold temp. 250-300 C for low conc.(<=10 ng) sample analysis.	23.AGC target value too high (>20,000), adjust to the range.	24.Once aliphatic sample involved, methane is not a proper reagent gas.	25.Ion neutralization can be sever if a H donor/ acceptor type of sample involved, set target val. at/ below 20,000 with emission 20 UA or more.	26.Check scan function line by line. Specifically, r.f./axial modulation amplitude and frequency, BBISO design, time intervals, etc.	27.Check scan function line by line. Specifically, r.f./axial modulation amplitude and frequency, BBISO design, time intervals, etc.	28.Check scan function line by line. Specifically, r.f./axial modulation amplitude and frequency, BBISO design, time intervals, etc.	29.AGC target val. small (<5,000), adjust to	30.Filament I small (<20 UA), adjust to range.	31.Electron multiplier GAIN (10^5) drop off.	32.Incorrect AM amlitude.	33.Carrier gas high(>2.0), adjust to 0.5-2.0	34.Carrier gas too high (>2.0), adjust to range.	35.Unstable carrier gas, controller/regulator.
				T	T													T		
				f	f															
				T																
				T	T															
					T							F								
						T	T		T	T	T	T	T	T	T					
									T		T	T	T							
									T											
										T										
								F			T			F						
											T				F					
	T																	T	T	
T																		T	T	T
			T			T	T													
T	T	T				t	t										F			
			F			F														
																				F
T	T	T	T				T	T								T	T		T	T
											F					T				
		T																		
													t							

as		Sample			Comu.			
36.Once oxygen peak presence, confirmed poor carrier gas quality.	37.Dirty plumbing, replace Cu pipes.	38.Sample absorbed at active surface in inj./col.	39.Complex sample matrix without pretreatment, using SFE, SPE to eliminate matrix effect.	40.Below detection limit when analying pesticides /HOCs, use split mode/combine ECD to confirm presence of trace analytes in sample.	41.Detach inst./check to locate problem source.	42.Check: power off/fuse blown; sample inj. to wrong col.; syr. defective; oven not heated/col. T low; no carrier gas flow; col. broken; inj. T low.	43.Possible failure: IEEE serial commu. board; IEEE cable; SAP PC board.	44.Possible failure: GC serial commu. PC board; GC CPU board; RS-232 cable; COM1 port.
T	T		T					
	T							
T	F							
F	f							
			F					
			T					
		T						
		T						
		T						
				T	T	T		
					T	F		
				T				
							T	
								T

Figure 5.5 The Knowledge Domain Matrix (KDM) used in the GCMSdiagnosis program

Lower case letters in the KDM indicate an 'OR' connection, a logical value implies either A or B plus a number of additional symptoms that can result from a given situation.

format. As one of the most general formats used in the expression of logical knowledge, the rule format consists of a set of conditional statements that uses IF condition(s) THEN conclusion(s) sentences to represent the heuristic knowledge. A tool kit, Rule-Editor, that can be used to perform batch conversion of the knowledge in a KDM form to rules has been developed in our laboratory to convert automatically the knowledge encoded in a KDM into a knowledge base file of rules (KBF). Table 5.2 illustrates part of the KBF file used by GCMSdiagnosis and gives an explanation on each of the sections, the key words and the delimiters. Leaking, Injection, Column, IonTrap, CarrierGas, Sample, and Communication are goal variables that represent the knowledge sub-domains in the KBF file. Refer to Appendix 5 for the complete current version of the KBF file used by GCMSdiagnosis. This particular rule (no. 20) encodes two layers of heuristics. At first, the rule asks for problem symptoms. In this case, if a high background noise is observed with a severely shortened ionization time and there are difficulties in tuning the RF voltage, these symptoms lead the expert system to conclude that the ion-trap is contaminated. At the next level, the system continues to ask whether conditioning the suspected part could effectively reduce the problem, and if not, the system then concludes that the ion-trap must be severely contaminated and physical cleaning of the device is required. In this way, a system user can be either told by the program to conditioning the ion-trap or informed to clean the device if it is necessary to solve the problem.

5.3.3 Implementation

An easy-to-use user interface is one of the key factors for the overall success of an expert system application. As indicated in a recent paper [32], the purpose of a knowledge-based system is to make available expertise that is beyond that of the user. The confidence in a system is a prerequisite for user acceptance of a program. No doubt, a more user friendly system will allow more efficient use of a program in problem-solving processes and will reduce the

Table 5.2 A rule file (KBF) example used by GCMSdiagnosis and brief explanations.

KBF Example	Explanation
Goal Section **FIND** Sample, Leaking, Injection, Column, CarrierGas, IonTrap, Communication; :	Goal Section starts with key word ***FIND***, ends with a delimiter ";". Goal variables are used to represent the knowledge subdomains.
Logic Expression Section RULE 20 **IF** Variable1=YES **AND** Variable8=YES **AND** Variable9=YES **AND** Variable7=NO **OR** Variable2=NO **THEN** IonTrap=Contamination severe/cleaning; **EXP** "When ionization time shortened severely and conditioning the device overnight ineffective in reducing the problem. It means that the ion-trap is badly contaminated and physical cleaning of the device is necessary."; :	This block begins with the keyword ***IF***, and the rule premises are connected by keywords ***AND***s and ***OR***s. The content after the keyword ***THEN*** is the conclusion of a rule block, and a delimiter ";" ends the rule block. An explanation of the conclusion is started by the keyword ***EXP***, and a delimiter ";" ends the explanation.
User Query Section ASK Variable1 "High background noise?" OPTION YES, NO; ASK Variable2 "Noise decrease as column temperature decrease?" OPTION YES, NO; ASK Variable7 "Conditioning suspected instrument part effectively reduce problem?" OPTION YES, NO; ASK Variable8 "Short ionization time (<15,000 ms)?" OPTION YES, NO; ASK Variable9 "RF tuning difficult" OPTION YES, NO; :	The ***ASK*** keyword begins a query block while a delimiter ";" ends it. Corresponding content of each variable is given by the ***ASK*** clause. And the ***OPTION*** part provides available answers to variables.

amount of time necessary to 'master' it.

In the GCMSdiagnosis project, we chose to develop our application within Microsoft Windows© 3.x environment because this provides excellent communication functionality through dynamic data exchange (DDE), object linking and embedding (OLE), dynamic linking library (DLL), and multiple-document interface (MDI). As shown in Figure 5.6, GCMSdiagnosis provides the user with a graphical user interface (GUI) developed in MS Visual Basic© (version 3.0, professional edition). The user interface incorporates the common features of MS Windows© based programs, such as a mouse, pull-down menus, and graphics, into the GCMSdiagnosis program. A user with little computer knowledge can easily become familiar with the operation of this program because the menus and icons provide an intuitive and self-explanatory operating environment.

The implementation results are discussed in the following. Figure 5.6 is a collection of screen-captures demonstrating the graphical features of the GCMSdiagnosis program. Figure 5.6(a) shows the common interface of GCMSdiagnosis, in which each of the different modular programs, such as described in Figure 5.3, are joined together as menu items. The next two screens are the screens of SPECview, the data handling module. Figure 5.6(b) displays a numerical window that is used to manipulate analytical data. In the picture, a data file from a GC-MS analysis is under review. Figure 5.6(c) exhibits the graphical features of the program in which the same data file opened in 5.6(b) is now displayed graphically. Three basic analytical signals are illustrated in 5.6(c) with the top showing a total ion chromatogram, the middle a single ion chromatogram, and the bottom a mass spectrum. The left portion of Figure 5.6(d) demonstrates a customized conversion efficiency (CE%) calculator that is used to calculate the conversion efficiency of an MSMS process. A number of CE% formulae have been stored in the program, such as the ones applicable to TCDDs (tertra- chlorodibenzo-dioxins) and TCDFs (tetrachlorodibenzo-furans).

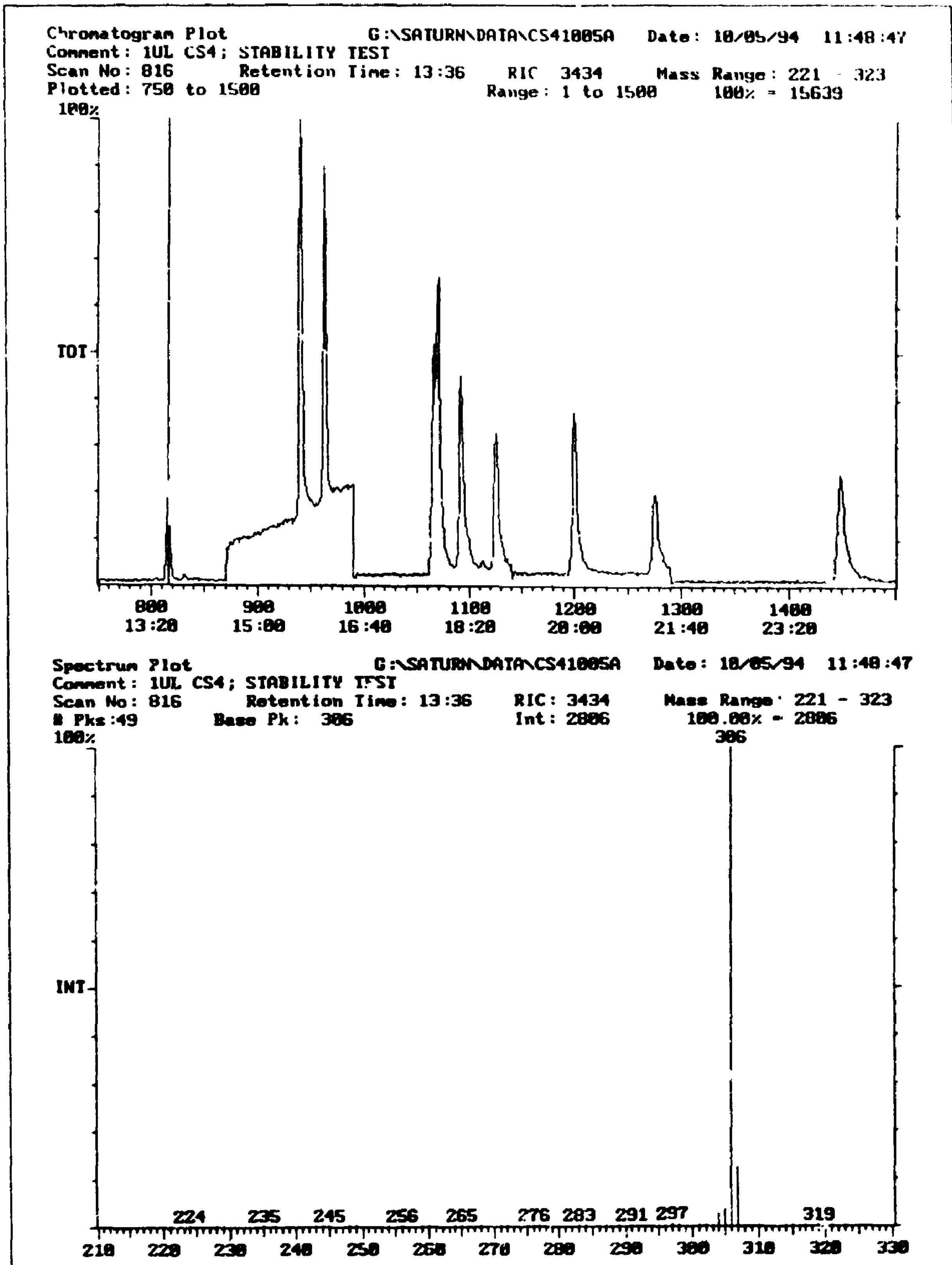
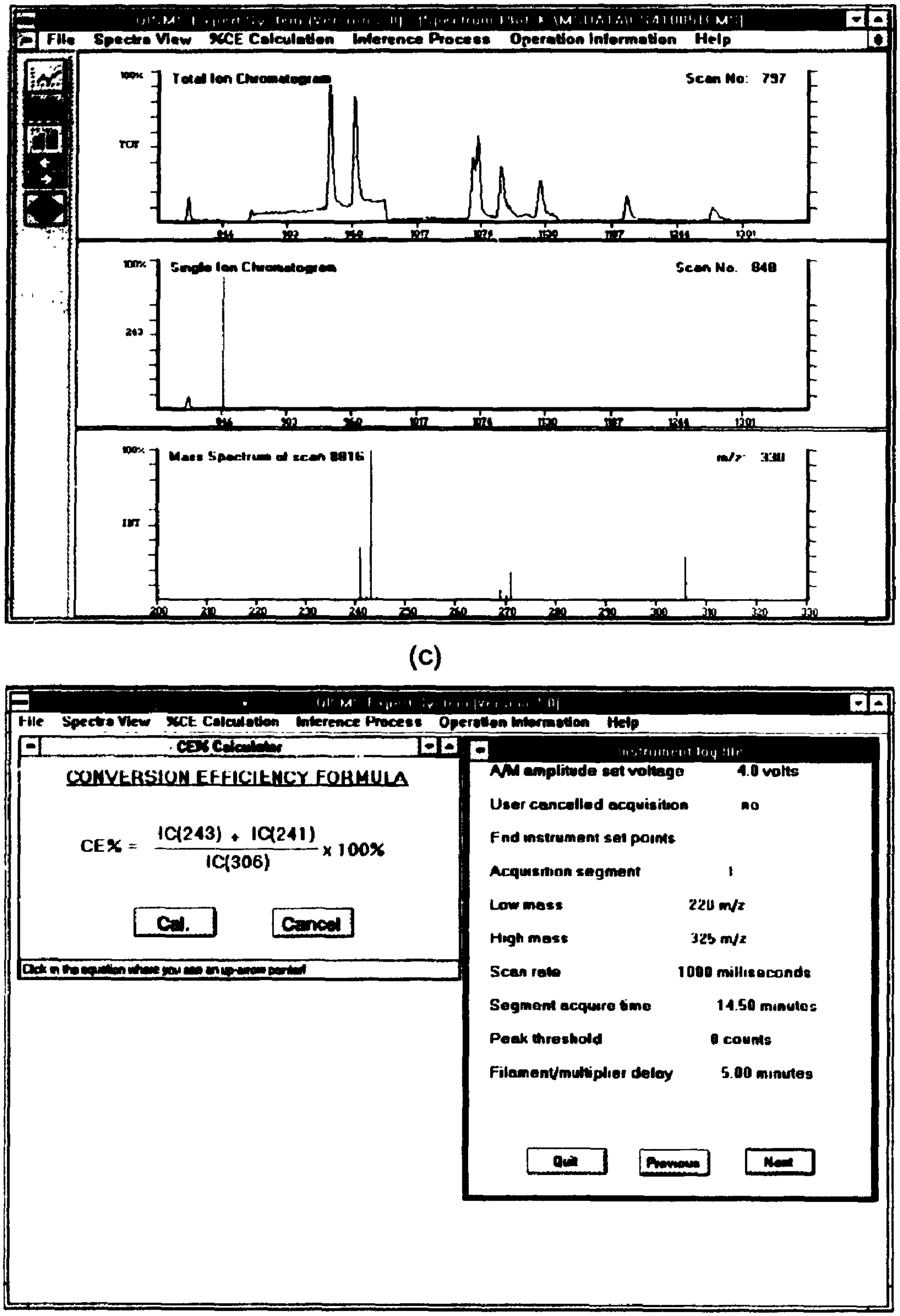


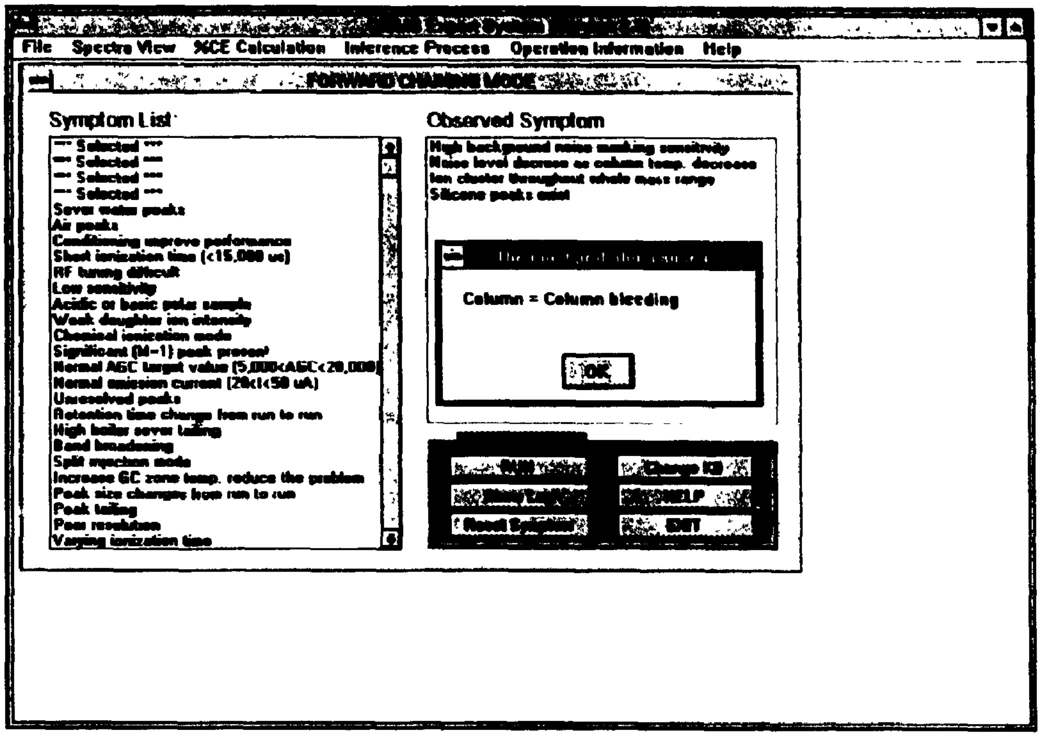
Figure 5.7 Signals pbtained from a SATURN™ III QISMS™ INSTRUMENT. The data show the signals from a GC-MSMS run of 1 μl of TCDF as the testing standard to examine the instrument condition. (a) is the TIC of parent ions, and (b) is the mass spectrum after applying a customized scan function for $[M+2]^+$ (m/z=306) ion isolation.



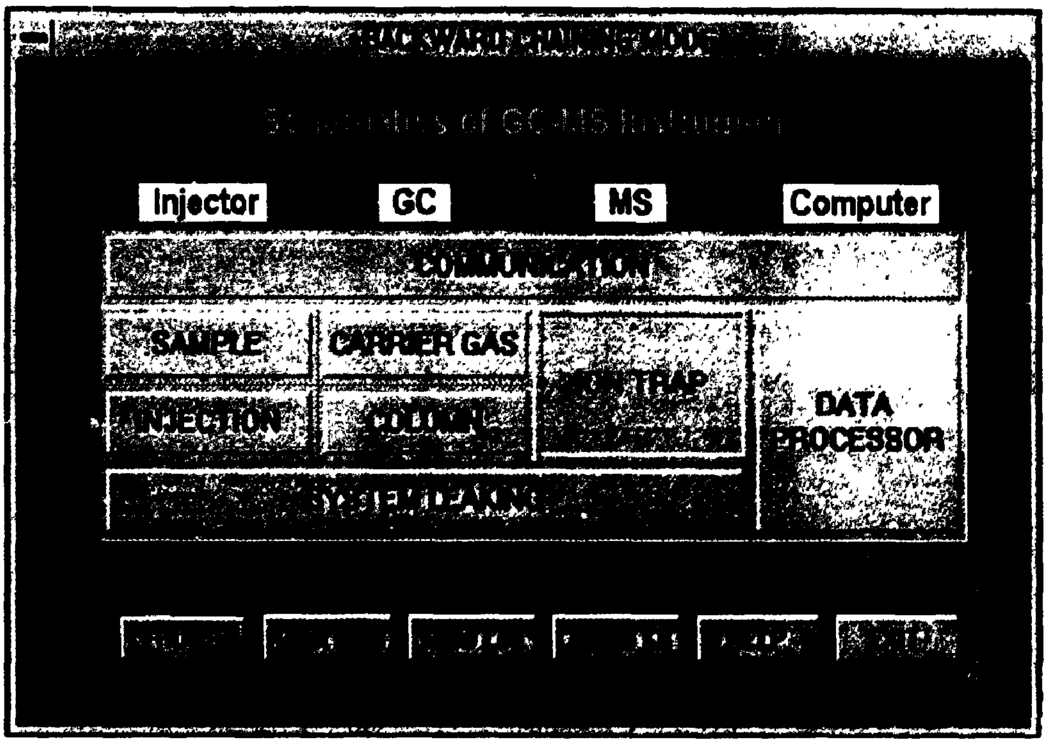
(c)

(d)

Figure 5.6 6(c) illustrates the graphical features of the program in which the same data file opened in 6(b) is now displayed graphically. 6(d) demonstrates a customized conversion efficiency (CE%) calculator.

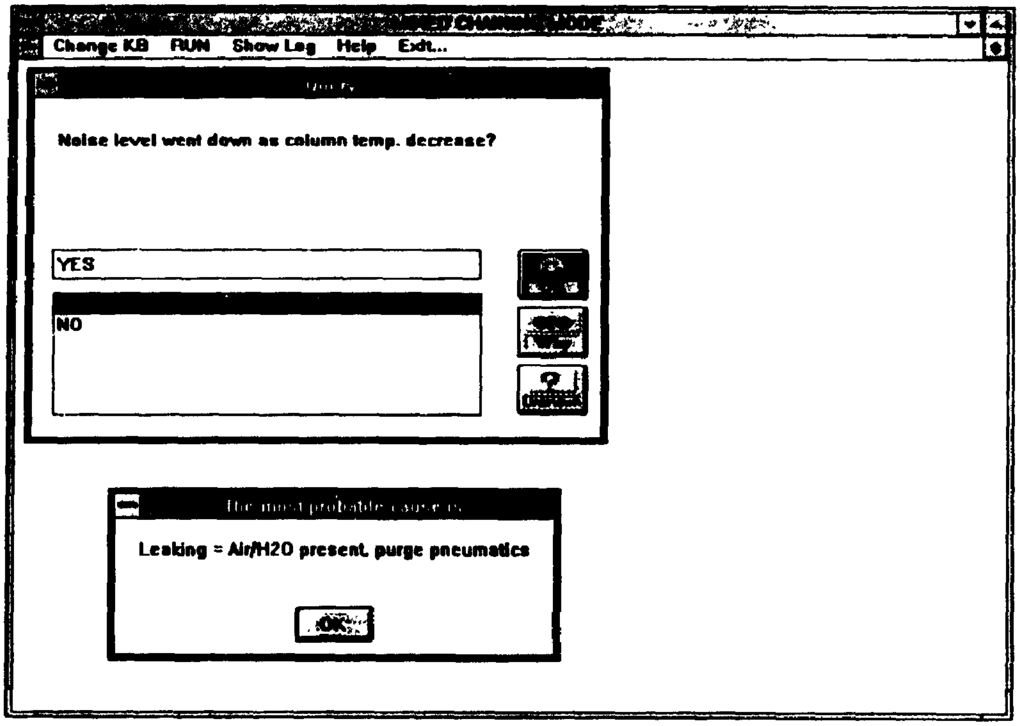


(e)

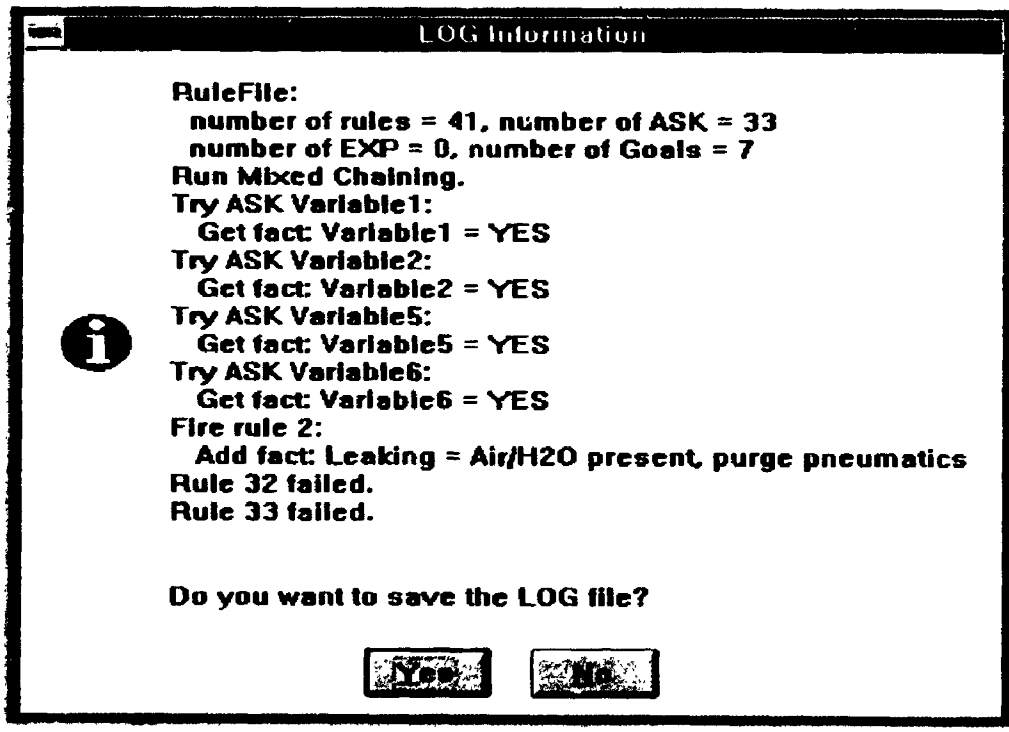


(f)

Figure 5.6 6(e) is the forward chaining inference window. (6f) shows an instrument schematics used for backward chaining inference strategy.



(g)



(h)

Figure 5.6 6(g) describes the mixed chaining inference window and a dialog box used for the system to display the inferential results. 6(h) shows a log file that explains the process the inference engine used to reach the current conclusion.

A template is also provided so that system users can specify their own formulae for various applications and save them for reuse. The right portion of Figure 5.6(d) displays a dialog box used to show instrumental parameters with which the analytical signals are generated. The next three screens show the three inference strategies incorporated in GCMSdiagnosis. Figure 5.6(e) is the window used for the forward chaining mode. Forward chaining, also called data-driven reasoning, runs successively by examining the premise of rules and fires those that are complete. Figure 5.6(f) is the window for the backward chaining. Backward chaining, or goal-directed reasoning, begins with a suspected goal and extends backwards through the knowledge base in an attempt to find evidence that could support the hypothesis selected by the user. Figure 5.6(g) is the window used for the mixed chaining mode This inference strategy refers to the process that uses both forward and backward reasoning with a single knowledge base. It starts in the forward chaining mode and when a situation arises for which insufficient evidence is encountered to support a hypothesis, the inference engine automatically switches to backward chaining mode and additional support for the hypothesis may be found by asking more goal-driven questions. Figure 5.6(g) also illustrates, in the bottom, the query window that the EAengine uses to present user questions and the answer box for displaying the conclusion(s) it reaches at the end of each inference process. Finally in Figure 5.6(h), a inference log file is displayed showing the actual route that the inference engine has taken to reach a particular conclusion. It is imperative to state that the knowledge base used by GCMSdiagnosis is separated from the main controlling program as an external library, and can be loaded by an on-line function call. In Windows programming, such a library is referred to as a dynamic linked library (DLL). The control program is able to use different knowledge bases for different applications. It is, therefore, possible for us to modify and/or update the knowledge base without touching the main computer code.

5.3.4 Example session with GCMSdiagnosis

In this section, we discuss an example of the use of the present version of the GCMSdiagnosis expert system. Figure 5.7 shows the analytical signals obtained from a SATURN™ III QISMS™ instrument. The instrument was recently set up in the MOEE laboratory as a testing site for both the SATURN™ III quadrupole ion-trap tandem mass spectrometer and the control software (SATURN™ Revision C). The experiment was designed to use 1μl of a mixture of tetra-octa chlorinated dibenzo-furans as the testing standard to examine the instrument conditions. The scan function was designed for isolation of tetrachloro-dibenzo-furan in the mixture of tetra-octa chlorinated furans and the subsequent MS/MS process in the electron ionization (EI) mode. As shown in Figure 5.7, part (a) is the total ion chromatogram (TIC) of the parent ions, and part (b) is the mass spectrum after applying a customized scan function for $[M+2]^{+\bullet}$ (m/z=306) ion isolation. Part (c) is the TIC ions, and part (d) is the corresponding mass spectrum of the isolated parent ions (m/z=306) after a CAD process.

Based upon the data acquired, GCMSdiagnosis was used to check first the quality of the analytical signals. For this purpose, the DIAGplatform module was launched and immediately an abnormal conversion efficiency (CE%) with a value of only 30.6% was reported by the program. According to the threshold value of the target compound of no less than 80%, GCMSdiagnosis concluded that there was a potential problem. Before we use GCMSdiagnosis to diagnose the possible cause(s), let us take a look at the data. The data shown in Figure 5.7 presented several obvious problematic symptoms, such as (1) After the MSMS process, the parent ion (TCDF at m/z 306) was still observed. (2) Compare to that of the parent ion, the TIC intensity of the daughter ions decreased significantly (>50% [illegible]) The late eluting components show an increased peak tailing. For an experienced analytical chemist, the potential cause for the problem could be a poorly designed isolation and CAD scan

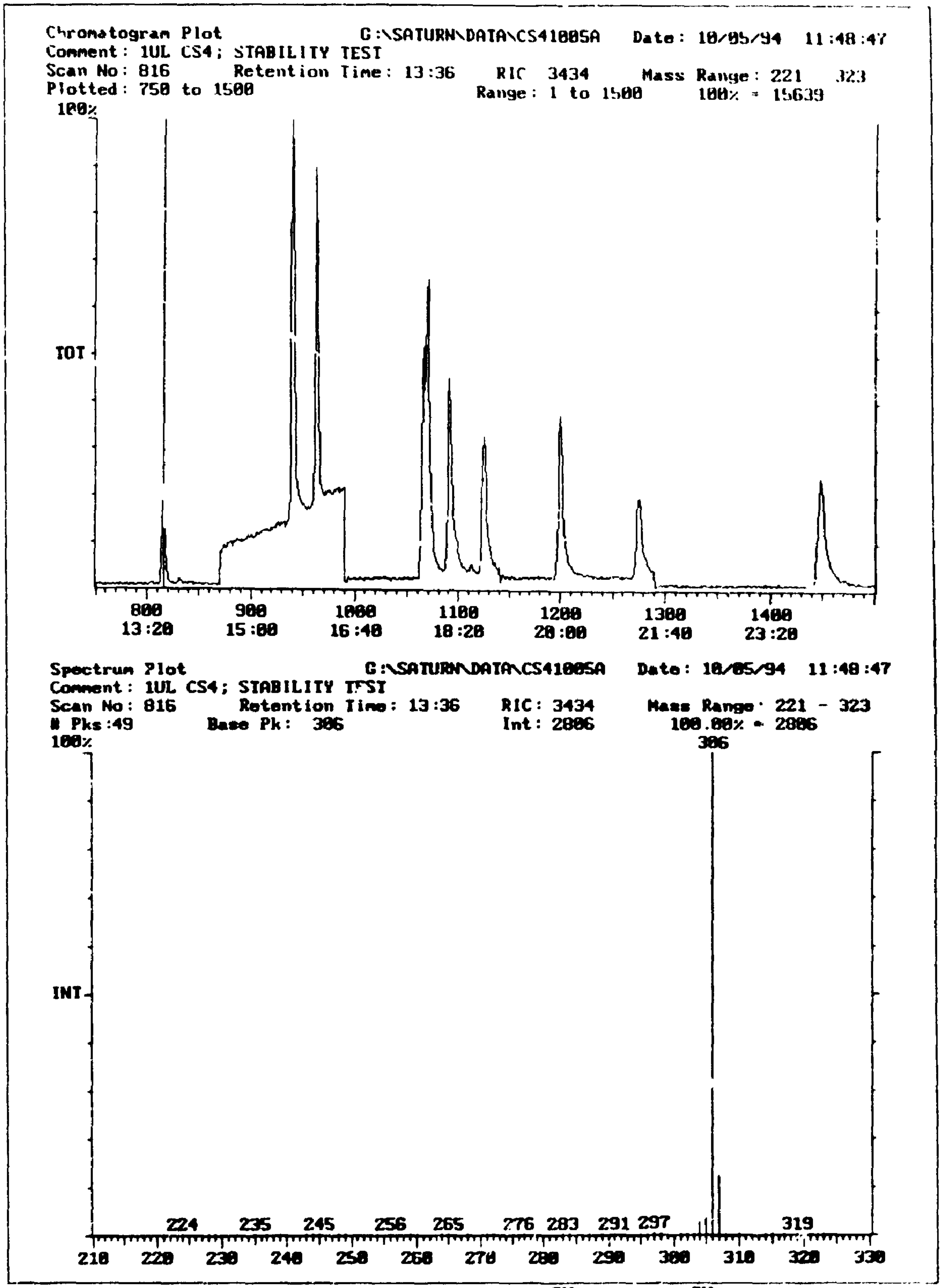


Figure 5.7 Signals pbtained from a SATURN™ III QISMS™ INSTRUMENT. The data show the signals from a GC-MSMS run of 1 μl of TCDF as the testing standard to examine the instrument condition. (a) is the RIC of parent ions, and (b) is the mass spectrum after applying a customized scan function for $[M+2]^+$ (m/z=306) ion isolation

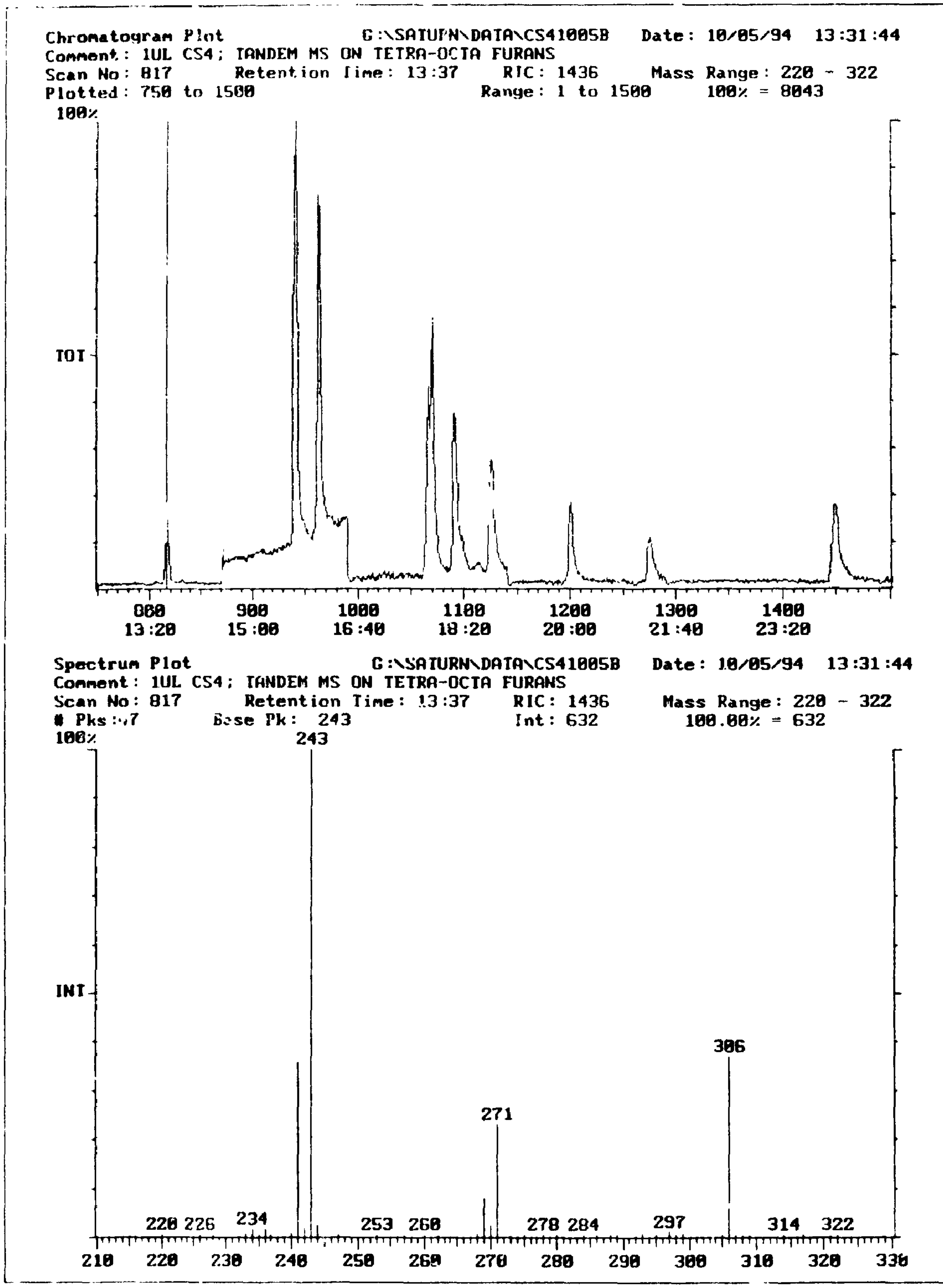


Figure 5.7 (c) is the TIC of daughter ions. and (d) is the corresponding mass spectrum of the isolated ions after a CAD process.

function, which could cause all but the last symptom mentioned above The data also indicated an improper GC temperature program Because the analysis was focused on TCDF, a relative light eluting component in the mixture of chlorinated tetra-octa furans, the tailing problem was not the major concern in this case.

Next, the diagnostic module (QISMSdiagnosis) was used to detect the possible causes for the low CE%. Different inferential processes were used in the diagnosis and the results are illustrated in Table 5.3. The questions that the system asked are listed in the first column. Columns 2, 3 and 4 show a list of answers to the questions (Q&As) using the mixed chaining inference mode Column 5 and 6 show the same list using the backward chaining inference mode, and column 7 and 8 are the results from using the forward chaining inference process. 'YES' to the question means a positive confirmation while 'NO' is a negative confirmation, and '?' means no confirmation, either 'NOT SURE', 'NOT OBSERVED' (but maybe exists), or 'DO NOT KNOW'. '--' indicates these questions were not presented in the inference process. Testing with the program has shown that a consistent answer was obtained despite the fact that we deliberately gave slightly different answers to the questions in order to reflect the diversity and differences in experience. The fastest and simplest approach is to use the forward chaining mode, however, such a process requires precise and accurate identification of problem symptoms that may not always be obvious to an operator with different levels of experience. Backward chaining offers a moderate approach that requires the operator to have a pretty good guess about which part of the instrument may be the source of an observed problem For the least experienced operator, the mixed chaining mode provides an interactive and directive approach because the system is able to lead the user through the diagnostic process by asking questions that are most relevant to the problem based on the heuristic knowledge encoded. The way that the system proposes questions in an inferential session is based on the statistical analysis of the knowledge represented in the knowledge base, therefore, it is able to use the

Table 5.3 Results from the use of the GCMSdiagnosis with test cases. Three inferential strategies are used with different answers given to the system. The test results show a consistent diagnosis of the problems.

Question List	Mixed chaining mode			Backward chaining		Forward chaining	
high background noise masking sensitivity	No	No	?	No	?		
noise level decrease as col temp. decrease	No	?	?	No	?		
conditioning suspected inst. component effective	?	--	--	?	--		
low sensitivity	Yes	Yes	Yes	Yes	Yes	√	√
severe water peak (H18/H19 << 10:1)	No	No	No	--	--		
retention time changing from run to run	No	No	No	--	--		
split injection mode	No	No	No	--	--		
peak size change from run to run	No	No	No	--	--		
high boiler severe tailing	Yes	?	?	No	Yes	√	
band broadening	Yes	?	Yes	Yes	--	√	
increase GC zone temp. reduce the problem	?	--	?	?	?		
poor resolution	No	--	No	?	?		
acidic or basic polar samples	No	No	No	--	--		
weak daughter ion intensity	Yes	Yes	Yes	Yes	Yes	√	√
chemical ionization mode	No	No	No	No	No		
(M+1) peak significant	--	--	--	No	No		
normal AGC target value (5,000 to 20,000)	Yes	?	?	?	?		
regular emission current (20 to 50 uA)	Yes	--	--	--	--		
RF tuning difficult	--	No	?	?	?		
Left over parent ion after tandem MS process	--	--	Yes	Yes	Yes	√	√
Total Questions asked or Facts supplied	**17**	**14**	**17**	**14**	**12**	**5**	**3**
Fired rule No.*	**26**	**27**	**28**	**28**	**28**	**28**	**28**

* Rule No. 26, 27, and 28 are parallel cases in which the conclusions indicate the same cause for the problems observed. Reference to Appendix 5 for details.

expertise to 'direct' a system user to look into certain symptoms specifically that the program 'thinks' may be key to a solution [22].

The paragraph above described a simplified example of using the program. However, because the answers suggested by the system are within the possibility and constraints specified by the user, it should be possible to achieve this kind of result in practice. The research phase of this project has been completed, and a compiled version of the GCMSdiagnosis program has been released It is now in the testing phase by a number of operational sites that have the facilities available We hope that testing with real cases will bring about modification and expansion of the current knowledge base As the knowledge base becomes more complete and precise, the system itself will be more reliable and practical for use.

5.4 CONCLUSIONS

Through the use of artificial intelligent technology, such as knowledge-based expert systems, scientists can be spared the most time-intensive steps in their routine analytical processes, and concentrate on at stages when they can work most proficiently, such as in interpreting research data and exploiting the meaning of the work. The development of expert systems will not only enhance novice instrument operators' performance in routine analyses to a higher level (expert-level), which would otherwise require extensive training and time, but will also benefit experts themselves.

In this chapter, we demonstrated a successful implementation of the KDM knowledge encoding scheme, however this time, the development work is in the analytical chemistry domain. As result, a diagnostic expert system for the GC-MS instrument has been developed. The GCMSdiagnosis program can be used with the computer that controls the Varian GC/QISMS™ instrument, and therefore, an on-line error detection and diagnosis is achieved Because the

SATURN software is a DOS-based program, the multi-tasking ability of GCMSdiagnosis inherent from MS Windows has not been fully used. The GCMSdiagnosis prototype has been developed based on the Varian GC/QISMS™ (GC-MS) system. However, the program should be readily amenable to other ion-trap GC-MS systems once the instrumental specifications are taken into consideration because the separation of the knowledge base from the controlling computer code is the intrinsic characteristic of a knowledge-based system.

Dedicated computers and programs are now commonly used for controlling analytical instrumentation. For example, diagnostic and self-tuning programs are provided in Varian SATURN™ systems. However, such computers and programs are not flexible and are not able to be customized to accumulate operational experience and expertise. As a result, we are often confronted with problems such as the one described in the example session. Those operational problems are not atypical, yet self-equipped diagnostic facilities usually are not able to detect them. In this regard, on-line knowledge based diagnostic programs will be useful during analytical processes since they contain domain specific knowledge that provides a supplement to the controlling software. Expert system technology provides the possibility of capturing operational knowledge by a computer program that can be used in dealing with similar cases or helping preserve and spread an expert's expertise.

The development of a diagnostic expert system consists of three basic steps. First, the relevant knowledge needs to be collected, structured, and represented in specific format to form the knowledge base. Then the knowledge base must be implemented in the computer system, and in the third step, the application has to be tested so that the expertise possessed by the program (stored in the knowledge base) can be refined and extended. Among them, the knowledge acquisition step is critical in the development cycle. It is important that knowledge acquisition have a focused, well-defined domain to start with.

Through this work, we found that it is not necessary to implement complete knowledge of the problem domain, and knowledge fragments represented in simple logical expressions, such as in rules, can provide very effective results. In this project, we have successfully developed a diagnostic expert system prototype for an ion-trap GC-MS instrument, yet the methodology and the approaches implemented in the development of this expert system prototype remain at the core of this research. Through several studies that have implemented the KDM knowledge encoding scheme for development of expert systems in the analytical chemistry domain, we conclude that the KDM scheme is effective and can be easily adopted for different situations. Verifying the general applicability of the KDM methodology was one of the tasks within our research.

5.6 REFERENCES

1. March, R. *Int. J. Mass Spectrom. Ion Processes*, **1992**, 118/119, 71-135.
2. Kaiser, R. E., Jr.; Cooks, R. G.; Stafford, G. C., Jr.; Syka, J. E. P.; Hemberger, P. H. *Int. J. Mass Spectrom. Ion Processes*, **1991**, 106, 79-115.
3. Williams, J. D.; Cox, K.; Morand, K. L.; Cooks, R. G.; Julian, R. K.; Kaiser, R. E. *Proc. 39th ASMS Conf. Mass Spectrom. Allied Topics*, **1991**, 1481-1482.
4. Londry, F. A.; Wells, G. J.; March, R. E. *Rapid Commun. Mass Spectrom.*, **1993**, 7, 43-45.
5. Plomley, J. B.; Koester, C. J.; March, R. E. *Organic Mass Spectrom.*, **1994**, 29, 372-381.
6. Eichelberger, P. H.; Bellar, T. A.; Donnelly, J. P.; Budde, W. L. *J. Chromatogr. Sci.*, **1990**, 28, 460-467.
7. Pereira, W. E.; Rosta, C. E.; Leiker, T. S. *Anal. Chim. Acta*, **1990**, 228, 69-75.
8. Hemberger, P. H.; Alaird, J. E.; Cameron, D.; Le nan, C. P.; Cannon, T. H.; Wolf, M. A., Kaiser, R. E. *Int. J. Mass Spectrom. Ion Progresses*, **1991**, 106, 299-306.
9. Williams, J. D.; Cox, K. A., Cooks, R. G.; McLuckey, S. A.; Hart, K. J.; Goeringer, D. E. *Anal. Chem.*, **1994,** 66, 725-729.
10. Guan, S. H.; Marshall, A. G. *J. Amer. Soc. Mass Spectrom.*, **1994**, 5, 64-71.
11. Eades, D. M.; Johnson, J. V.; Yost, R. A. *J. Amer. Soc. Mass Spectrom.*, **1993**, 4, 917-929.
12. Zerega, Y.; Andre, J.; Brincourt, G.; Catella, R. *Int. J. Mass Spectrom. Ion Processes*, **1994**, 132, 57-65.
13. Zerega, Y.; Andre, J.; Brincourt, G.; Catella, R. *Int. J. Mass Spectrom. Ion Processes*, **1994**, 132, 67-72.

14. Zerega, Y.; Andre, J.; Brincourt, G.; Catella, R. *Int. J. Mass Spectrom. Ion Processes*, **1994**, 135, 155-164.

15. Eckenrode, B. A.; McLuckey, S. A.; Glish, G. L. *Int. J. Mass Spectrom Ion Processes*, **1991**, 106, 137-142.

16. Orht, R. G.; Hail, D.; Hileman, F. D.; Weber-Grabau, M., Kelley, P. E. *EPA /APCA Int. Symp: Meas. Toxic and Related Air Pollutants*, **1989**, p. 291.

17. Plomley, J. B.; Koester, C. J.; March, R. E. *Anal. Chem.*, **1994**, 66, 4437-4443.

18. Burlingame, A. L.; Boyd, R. K.; Gaskell, S. J. *Anal. Chem.*, **1994**, 66, 634R-683R.

19. Varmuza, K. *Int. J. Mass Spectrom. Ion Processes*, **1992**, 118/119, 811-823.

20. Fekih-Ahmed, F.; Faltings, B. V.; Bachmann, C. *Chemometr Intel. Lab Syst.*, **1993**, 21, 267-273.

21. Brandt, J.; Hitzmann, B. *Anal. Chim. Acta*, **1994**, 291, 29-40.

22. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 956-968.

23. SATURN™ - Gas Chromatography Mass Spectrometer Operator's Manual, 03-914350-00-A, Varian Associates, Inc., Walnuit Creek, CA , USA.

24. Du, H.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 33-41.

25. Du, H.; Larhiri, S.; Huang, G., Stillman, M. J. *Anal. Chim Acta*, **1994**, 296, 21-31.

26. March, R. E.; Hughes, R. J. *Quadrupole Storage Mass Spectrometry, Chemical Analysis Series, VOl. 102*, Wiley & Sons, New York, 1989.

27. Huang, G.; Stillman, M. J. An Expert System Shell Developed for Use in Microsoft Windows™ Environment at MJS Laboratory. Unpublished results.

28. Zhu, Q.; Stillman, M. J. *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 945-955

29. Zhu, Q.; Stillman, M. J. Expert Systems and Analytical Chemistry. Recent Progress in the ACexpert Project. *J. Chem. Inf. Comput. Sci*, **1996**, in press.

30 Lahiri, S ; Stillman, M. J. *Anal. Chem.*, **1992**, 64, 283A-291A.

31. Lahiri, S., Yuan, B., Stillman, M. J. *Anal. Chem.*, **1994**, 66, 2954-2963.

32. Hochman, Z.; Pearson, C. J.; Litchfield, R. W. *Agricultural Systems*, **1994**, 44, 217-235.

CHAPTER 6
CONCLUSIONS

6.0 GENERAL SUMMARY

What analytical chemistry tries to obtain through experiments is information from material systems or objects by chemical, physical, or physico-chemical techniques [1]. These techniques are often sophisticated, very precise and sensitive. Although most of these techniques are scientifically and technologically sound and well founded, it is not possible to predict theoretically which method will give the best answer(s) to a specific problem. Therefore, there is no single scientific procedure to be followed to accomplish an entire analytical process from problem to solution. This absence of theoretical, predictable routines for use of analytical procedures can be compensated for by using rule-of-thumb based on past experience. Hence, chemical analyses depend on highly trained and experienced chemists. However, such rule-of-thumb does not have general applicability and only experience can tell whether to apply it in a given situation. This is the reason that the integration of AI technology, such as the expert system, has become increasingly commonplace for laboratories.

Expert system technology offers the possibility of combining theories with heuristic expertise that field expert(s) have acquired through years of practice. These systems, therefore, can have the flexibility to cope with complex problems of the real world that otherwise may need the presence of human expert(s) [2]. Expert systems will not likely replace human experts, instead, they offer help by releasing experts from their day-to-day routine work and enabling them focus on the innovative part of their job. Expert systems also provide a way to document heuristic knowledge and help junior level technicians to perform their duties at a level that otherwise may require the assistance from human experts

The differences between conventional programs and expert systems are briefly summarized in Table 6.1. In reality however, such distinction may not be as clear as Table 6.1 might suggest. For example, some expert systems use conventional mathematical probabilistic methods to deal with uncertainty, and a complex expert system may contain a series of embedded procedural programs.

Table 6.1 Comparison of expert systems with conventional programs

Expert Systems	Conventional Programs
Represent and use knowledge	Represent and use data
Heuristic	Algorithmic
Use inferential process	Iterate repetitively and exhaustively
May have numerous information pathways	Use a decision sequence with all decision rules preprogrammed in routines
Allow for missing information	Stop if data missing
Derive best decisions possible	Derive one solution consistently
Can also use concepts	Use numbers and equations
Provide uncertain results ranked by likelihood	Present a result with confidence limit
Appropriate for imprecise information and conceptual problems	Appropriate for accepted ways to solve a problem, especially with only one solution

The implementation of computer technology in the field of analytical chemistry came in three steps [3]. Computers were initially used in instruments for process control and data logging. Applications soon advanced to the next step in which computers were used for data processing and quality assurance purposes. For example, the implementation of the Fourier-Transform algorithm into spectroscopy has brought forth the development of the new field of measuring technology. Now, computer technology is incorporated into areas in which extensive expertise is involved. This step has been made possible through the implementation of AI techniques, such as is found in expert systems, pattern recognition, neural networks, and machine adaptation and learning. However, the biggest challenge to the widespread implementation of expert system technology into scientific applications comes from translation of heuristic knowledge and established algorithms into a format accessible to the computer program algorithm.

6.1 PROJECT SUMMARY

This thesis started with a brief review of the recent advances made in application of expert system technology in the chemistry domain. While this thesis exclusively summarized the results from the MJS group in implementation of the ACexpert project, it has focused on representing the results from areas in which I am primarily involved. We believe that expert system technology will be more fully integrated into the chemistry domain in the future, with applications in analytical chemistry leading the way. Examples of small, 'light' expert systems, such as those developed in the ACexpert project, are expected to be found soon on computers on laboratory benches

6.1.1 Encoding chemical knowledge

One of the most important features of an expert system is that it can be improved by the addition of further knowledge to the system. Provision should be made for maintaining the knowledge base. Because rules do overlap and complex interrelationships exist, maintenance of the knowledge base is indeed a complex and difficult task. On the other hand, the knowledge base of an expert system should be carefully defined according to the role it is designed to play. For example, if an expert system is to be used to screen possible analytical protocols and select the ones that are most appropriate for a user-defined situation, the knowledge base of this system should be broad without great depth. In contrast, if the system is to function as an expert in performing a specific, yet complex analysis, the knowledge base should rather be limited to a well-defined field of expertise with great possible depth.

Knowledge Domain Matrix is a knowledge encoding alternative that has been proved to be effective in coding chemical knowledge through the development of a number of applications in the ACexpert project. The KDM scheme provides a well-defined causal knowledge model (see Figure 5.4) that a knowledge engineer could follow closely to interpret domain expertise in terms of computer data structure. The modification of a knowledge base developed using the KDM scheme is flexible and intuitive because, unlike a tree-structure, the logical relations in the KDM format are not hard-wired and can be edited without affecting the rest. In addition, it provides the visual accessibility to the chosen knowledge domain by encoding it in a knowledge primitive (a matrix) upon which the knowledge base can be derived. Therefore, knowledge engineers do not need to memorize the logical connections and can make changes forward and backward easily.

6.1.2 Design of the user interface

The success of knowledge-based expert systems depends to a large extent on the man-machine interface, both during development when the domain expert and knowledge engineer provide the system with a representation of knowledge, and later when the end-user tries to implement the system. The learning curve for chemists to develop an expert system is fairly long. This is mainly due to their limited ability in understanding computing language(s) and subsequently writing program code, among which a larger portion is the code of the user interface.

We have developed the ACexpert graphical user interface within the Microsoft Windows© 3.x environment. Together with the KDM knowledge encoding scheme and the inference engine provided by EAshell©, these three key components form a rather loose expert system shell that can be reused for different knowledge bases. The test of the applicability and readiness of the implementation of this shell is currently being undertaken as a fourth year thesis project in which the shell is implemented in an effort to develop a diagnostic expert system for the Varian HPLC star instrument.

6.1.5 The SPILLexpert project

In the environmental field, the potential use of expert systems lies beyond their immediate use for emergency response to chemical spills. As more packages and systems become available, AI will be extended to more and various applications. There is a strong possibility of using expert systems collectively [4]. For example, SPILLexpert could be used in conjunction with environmental assessment systems to monitor the potential impacts of spilled chemicals and mitigation measures on the environment, or in conjunction with ecological models, geographical information systems, even computer-aided drafting and modeling packages to provide sophisticated trends and graphics of the environmental degradation behaviors of the released chemicals. Expert systems also have great potential for use in environmental training and education, especially in circumstances where there is a shortage of expertise, as in the field of chemical spill response.

Canada needs a more coherent national capability to respond to chemical spills. The current systems adopted by various public and private sectors need improvement to approach the establishment of a network system, in which the use of computer technology, specifically the use of knowledge-based expert systems will play an indispensable role. The SPILLexpert project is a step taken towards this goal. Results from the project reveal that a large volume of information regarding chemical spill response is available, yet there lacks an effective information retrieval system, more so is the lack of ability to compare case histories and reapply heuristics (experience) in planning of more precise and reliable countermeasures for a current situation. SPILLexpert is at present a very rudimentary expert system which requires further development, particularly the interactions with field personnel to adjust and modify it's knowledge base. It does, however, demonstrate the possibility of implementing expert systems for regulatory and procedurally-based matters, such as response to chemical spills, therefore providing a more flexible and readily-available source of expertise in a rapid expanding field. More likely, such systems will be integrated with other software that will enhance their application capabilities.

6.1.4 GCMSdiagnosis project

Failure detection, testing and maintenance of analytical instrumentation are knowledge intensive tasks that rely extensively on expertise [5]. Apart from using embedded testing and tuning procedures, skilled analytical chemists need to use heuristics and an understanding of how a system works to solve a real-world problem. The purpose of diagnostic expert systems in this area is to exploit heuristic knowledge in conjunction with instrumental failure detection to generate diagnostic procedures to ease operator's work-load.

The requirements for an ideal trouble-shooting system include minimizing the probability of false alarms and of not detecting failures, reducing time taken, effective integration and expansion of knowledge, and ease of instrument maintenance. The major tasks are to devise a knowledge representation scheme to allow failure events to be analyzed by the merging of diverse information, and to develop an intuitive user interface to grant easy utilization of such systems so that users (chemists) do not need to pay much attention in learning the operation and are able to use them effectively. The GCMSdiagnosis project is another exercise of developing diagnostic expert systems for analytical instrumentation in which previously gained experience and concepts [6,7] are implemented to result in a fully functional prototypic expert system [8].

6.2 THOUGHTS ON FUTURE WORKS

6.2.1 The merger with neural networks

There is a clear call for the capability to deal optimally with the affluence of data and explosion of knowledge occurring in any research field [9], for example, in the fields of analytical and environmental chemistry Computer-supported decision making may be the solution to this question. Since expert systems are built using knowledge from domain experts, they may perform more or less like human experts. More so, a computer-based expert system will not get tired and is always available. It is therefore more consistent and reliable than its human counterparts.

However, the use of expert systems is not without problems. First of all, unless the knowledge is clearly defined and well accepted, knowledge acquisition (i.e. what to put into the knowledge base and what to leave out, and more particularly how to obtain relevant knowledge from human expert(s) and represent it as a computerized data structure) will continuously be a problem. Next, the problem-solving ability of expert systems depend largely on their pre-programmed knowledge. Like their makers, expert systems are not necessarily infallible. Typical expert systems often do not include the ability to learn from errors and be self correcting.

Neural networks rely on the concept of modeling a computer system on the structure of the human brain. In a neural network, individual processing elements communicate via a rich set of interconnections with variable weights. Just like biological neurons, processing elements exist in a variety ways. One or more inputs are regulated by the connection weights to change the stimulation level within the processing element. The output of the processing element is related to its activation level and is non-linear or discontinuous in nature. Unlike expert systems, a neural network system is taught by training cases through adjustments of interconnection weights of the transfer functions of the element rather than pre-programmed. Thus the memories ('knowledge') are stored in the interconnection network as a pattern of weights. Information is processed as a spreading changing pattern of activity distributed across many elements. Perhaps the future of artificial intelligent systems may rely on a combination of the recognition ability of the neural network to arrive at the solution to a problem and the rationalization capability of the expert system to explain the reasons for the solution given.

6.2.2 Interface expert systems to LIMS

Good laboratory practice (GLP) was first established by the US Food and Drug Administration (FDA) in 1979 [10]. At the center of GLP is the implementation of laboratory information management systems (LIMS). A LIMS is a computer application concerned with the management of data and disseminating the information produced by modern laboratories [11]. The essence of a LIMS is a database and associated software modules that are tailored to the operation of individual laboratory or laboratories within an organization. A LIMS is intended to integrate sample information with results produced from analytical procedures carried out on it. Murphy has argued that a large portion of LIMS systems installed to date are merely laboratory data management systems [12]. The same article underlines the difference between data and information and the need to bring about such a conversion.

At present LIMS and chemometrics, such as expert systems, are considered very separate entities with little in common. However, they are complimentary and can merge to the mutual benefit of the future LIMS that can be capable of providing quality information in a timely manner. This opportunity is highlighted by Megargle when he states [13]: "Artificial intelligence is likely to have an impact on the LIMS of the future. Interpretations based on the combined results from different workstations will be performed within the LIMS. The knowledge to do this will be entered into artificial intelligence data structure... using the LIMS results to analyze a problem, draw conclusions, and make technical and business decisions."

6.3 REFERENCES

1 Kateman, G *Analyst*, **1990**, 115, 487-495.

2 Klaessens, J. W A , Kateman, G ; Vandeginste, G. M. *Trends Anal. Chem.*, **1985**, 114-122.

3. Kateman, G. *Trends Anal. Chem.*, **1984**, 3, VII.

4 Geraghty, P. J. *J. Environ. Manage.*, **1993**, 39, 27-38.

5. Pau, L. F. *Expert Syst.*, **1986**, 3, 100-112.

6 Lahiri, S.; Stillman, M J. *Anal. Chem.*, **1992**, 64, 283A-291A.

7. Du, H.; Lahiri, S., Huang, G.; Stillman, M. J. *Anal. Chim. Acta*, **1994**, 296, 21-31.

8. Zhu, Q., Du, H.; Plomley, J. B.; March, R. E.; Stillman, M. J. GCMSdiagnosis - An expert system for diagnosis of problem GC-MS data. *Anal. Chim. Acta*, **1996**, submitted.

9. Place, J. F ; Truchaud, A.; Ozawa, K.; Pardue, H., Schnipelsky, A. *Clin. Biochem.*, **1995**, 28, 373-389.

10. Stokes, T. *Chemometr. Intell. Lab. Syst.*, **1993**, 21, 131-138.

11. McDowall, R. D.; Leavens, W. J.; Massart, D. L. *Chemometr. Intell. Lab. Syst.*, **1992**, 13, 221-230.

12. Murphy, M. *Chemomtr. Intell. Lab. Syst.*, **1991**, 13, 37-41.

13. Megargle, R. *Anal. Chem.*, **1989**, 61, 612A-621A.

Appendix 1: ERexpert Factbase details (in MS ACCESS)

The design and implementation of the ERexpert Factbase using MS ACCESS is illustrated in Appendix 1. Figure A1-1 shows the list of the fact tables used in ERexpert module that hold the subjective information relevant in response to chemical spills. Figure A1-2 is a list of the query tables. Action/topic oriented information can be drawn from different fact tables dynamically to form query tables to present system user more complete information set. Figure A1-3 is a list of forms, and Figure A1-4 is a lists of reports. Both are used to present information (facts) in a more structured and readable way.

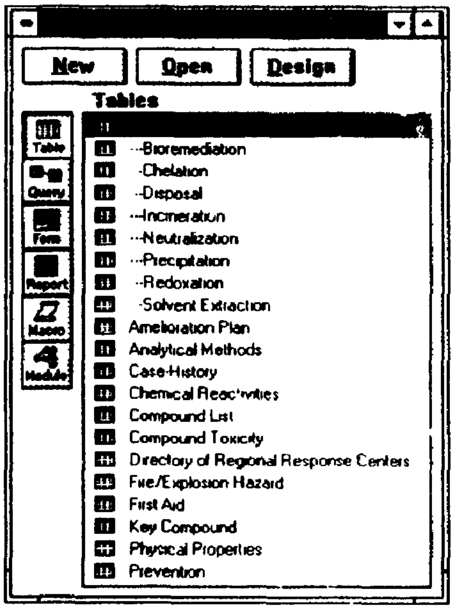


Figure A1-1

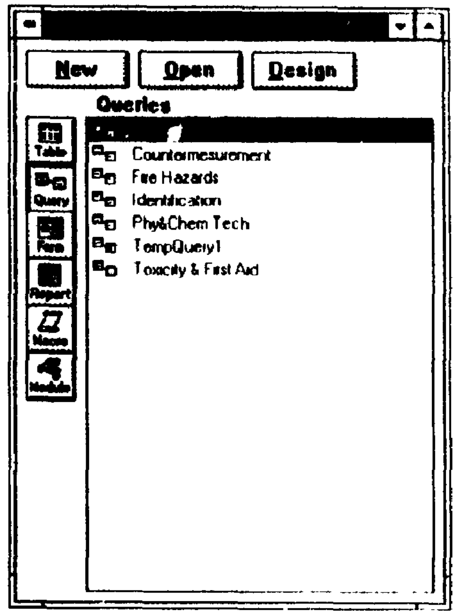


Figure A1-2

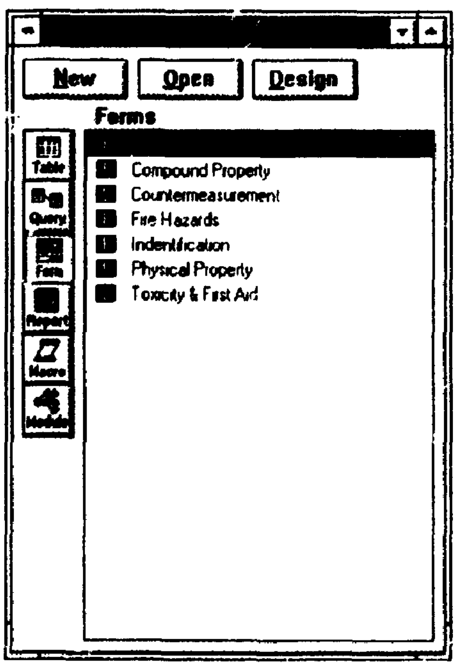


Figure A1-3

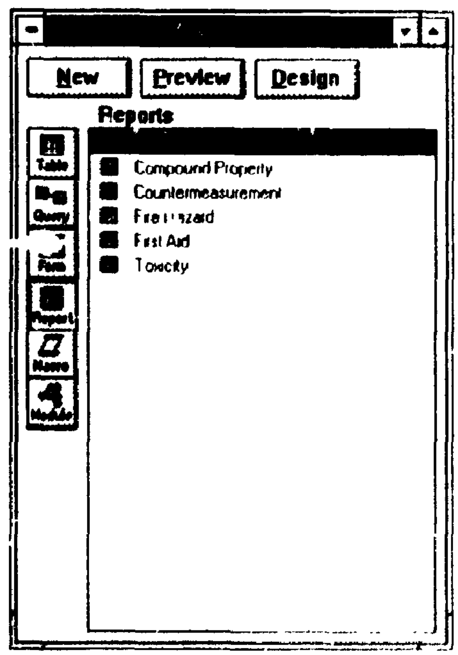


Figure A1-4

Table A1-1 exemplifies the first page of the "Compound List" table. This table is one of the fact tables in which the names, CAS registration numbers, Category IDs, and synonyms of a total of 396 compounds are listed.

Comp ID	Category ID	CAS No	Comp Name	Phase ID	Family	Synonyms
1	1A		Acetaldehyde			
2	1A		Acetone			
3	1A		Acetonitrile			
4	1A; 2B		Allyl chloride			
5	1A; 2E	71-43-2	Benzene	VAPOR,LIQUID	Aromatic compo	Benzol; Phenyl hydrid
6	1A; 2B		Butadiene, inhibited			
7	1A.		Butane			
8	1A; 2B; 2D		n-Butyl alcohol			
9	1A		sec-Butyl alcohol			
10	1A		tert-Butyl alcohol			
11	1A.		Butylene			
12	1A; 2B; 2D		n-Butyraldehyde			
13	1A.		Cyclohexane			
14	1..		Diethylamine			
15	1A; 2B		Diisobutylene			
16	1A		Dimethylamine			
17	1A		1,1-Dimethylhydrazine			
18	1A; 2B; 2E		Distillates: flashed fee			
19	1A; 2B; 2E		Distillates: straight ru			
20	1A; 2B; 2D		Ethyl acetate			
21	1A		Ethyl alcohol			
22	1A; 2B; 2E		Ethylbenzene			
23	1A.		Ethyl chloride			
24	1A		Ethyleneimine			
25	1A		Ethylene oxide			
26	1A; 2B; 2D		Ethyl ether			
27	1A; 2B; 2E		Gasoline blending stoc			
28	1A; 2B; 2E		Gasoline ble. ding sto:			
29	1A; 2B; 2E		Gasolines: automotive			
30	1A; 2B; 2E		Gasolines: aviation (<			
31	1A; 2B; 2E		Gasolines: casinghead			
32	1A; 2B; 2E		Gasolines: polymer			
33	1A; 2B; 2E		Gasolines: Straight ru			
34	1A; 2B; 2E		Heptane			
35	1A; 2B; 2E		1-Heptene			
36	1A; 2B; 2E	110-54-3	Hexane	VAPOR,LIQUID	Aliphatic hydroc	None
37	1A; 2B; 2E		1-Hexene			
38	1A		Hydrogen sulfide			
39	1A		Isobutane			
40	1A		Isobutylene			
41	1A; 2B; 2D		Isobutyraldehyde			
42	1A; 2B; 2E		Isohexane			
43	1A; 2B; 2E		Isopentane			
44	1A; 2B; 2E		Isoprene			
45	1A		Isopropyl alcohol			
46	1A		Liquefied petroleum g			
47	1A; 2B; 2D		Methyl acrylate			
48	1A		Methyl chloride			
49	1A		Methyl ethyl ketone			
50	1A, 2B, 2D		Methyl methacrylate			
51	1A, 2B; 2E		Naphtha. VM&P (75%			
52	1A		Nitromethane			
53	1A, 2B, 2E		Pentane			
54	1A; 2B, 2E		1-Pentene			
55	1A; 4A; 4C		Phosphorous pentasul			
56	1A		Propane			
57	1A; 1E		Propionaldehyde			
58	1A		Propylene			
59	1A; 2B; 2E		Toluene			
60	1A		Trimethylamine			

Table A1-1

Figure A1-5 shows the design of a query table, named Countermeasurement, which draws on relevant information from 4 individual fact tables. The design demonstrates the connections used to draw information from various table.

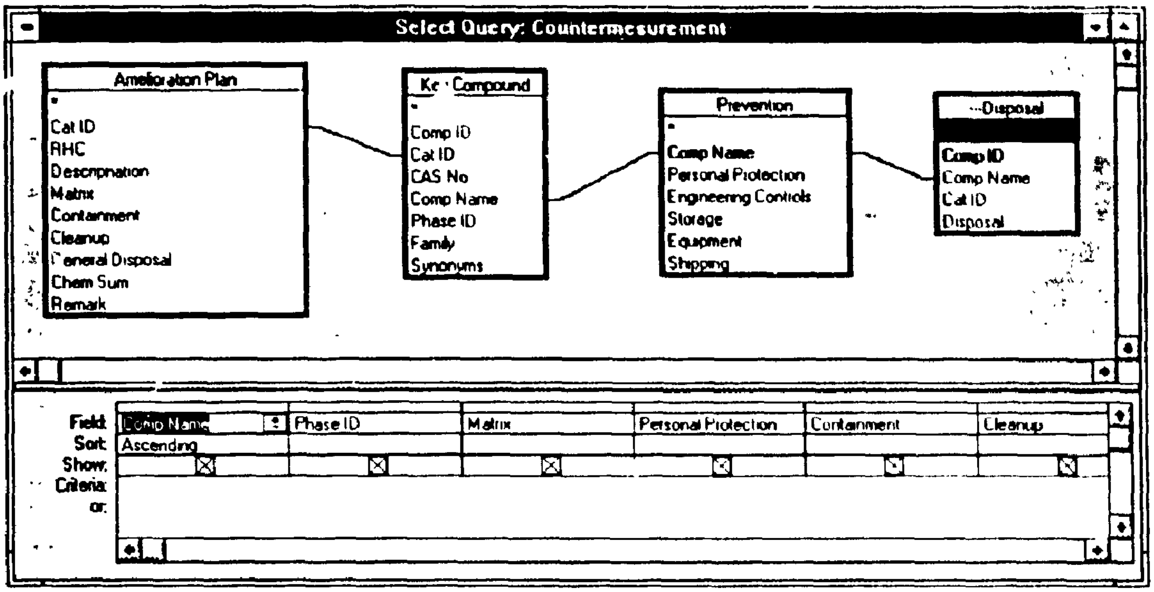


Figure A1-5

Table A1-2 is the actual display of this query table. However, the large amount of information held in cells is not readable. In the next two pages, we show two different designs that the program uses to display information in a more structured way. Figure A1-6 is a REPORT format in which the same information shown in Table A1-2 is now displayed. Figure A1-7 shows a FORM format.

Select Query: Countermeasurement

Comp Name	Category ID	Phase ID	Matrix	Personal Protection	Containment	Cleanup	Storage	Shipping	Disposal
Benzene	1A	VAPOR	On land	Wear appropriate cloth	Stop the leaking source i	1A1 No effective	Store in suital	DOT label is 'Fl	Incineration
Benzene	1A	VAPOR	Watercourse	Wear appropriate cloth	Stop the leaking source i	1A2 - No effective	Store in suital	DOT label is 'Fl	Incineration
Benzene	2E	LIQUID	On land	Wear appropriate cloth	Stop the leaking source.	2E1 - "Pump into s	Store in suital	DOT label is 'Fl	Incineration
Benzene	2E	LIQUID	Watercourse	Wear appropriate cloth	Stop the leaking source i	2E2 Using "Skimr	Store in suital	DOT label is 'Fl	Incineration
Bromine	3B	LIQUID	On land	Wear appropriate cloth	Stop the leaking source	3B1 "Pump into s	Store in suital	DOT label is 'Ci	Large volume
Bromine	3B	LIQUID	Watercourse	Wear appropriate cloth	"Underwater dike" "Barr	3B2 - "In-place che	Store in suital	DOT label is 'Ci	Large volume
Bromine	4E	LIQUID	On land	Wear appropriate cloth	Stop the leaking source i	4E1 - "Pump into s	Store in suital	DOT label is 'Ci	Large volume
Bromine	4E	LIQUID	Watercourse	Wear appropriate cloth	For water soluble chemic	4E2 - "Adsorption"	Store in suital	DOT label is 'Ci	Large volume
Hexane	1A	VAPOR	On land	Wear appropriate cloth	Stop the leaking source i	1A1 - No effective	Store in tightly	DOT label is 'Fl	Incineration
Hexane	1A	VAPOR	Watercourse	Wear appropriate cloth	Stop the leaking source i	1A2 - No effective	Store in tightly	DOT label is 'Fl	Incineration
Hexane	2B	LIQUID	On land	Wear appropriate cloth	Stop the leaking source.	2B1 "Pump into s	Store in tightly	DOT label is 'Fl	Incineration
Hexane	2B	LIQUID	Watercourse	Wear appropriate cloth	Stop the leaking source i	2B2 - "Skimmer" "	Store in tightly	DOT label is 'Fl	Incineration
Hexane	2E	LIQUID	On land	Wear appropriate cloth	Stop the leaking source.	2E1 "Pump into s	Store in tightly	DOT label is 'Fl	Incineration
Hexane	2E	LIQUID	Watercourse	Wear appropriate cloth	Stop the leaking source i	2E2 - Using "Skimr	Store in tightly	DOT label is 'Fl	Incineration
Hexane	2G	LIQUID	On land	Wear appropriate cloth	Stop the leaking source	2G1 "Pump into s	Store in tightly	DOT label is 'Fl	Incineration
Hexane	2G	LIQUID	Watercourse	Wear appropriate cloth	Stop the leaking source i	2G2 "Skimmer" "	Store in tightly	DOT label is 'Fl	Incineration
Phenol	1B	VAPOR	On land	Wear appropriate cloth	Stop the leaking source i	1B1 "Adsorption"	Store in suital	DOT label is 'Pr	Incineration
Phenol	1B	VAPOR	Watercourse	Wear appropriate cloth	Stop the leaking source i	1B2 - "Adsorption"	Store in suital	DOT label is 'Pr	Incineration
Phenol	1E	VAPOR	On land	Wear appropriate cloth	Stop the leaking source i	1E1 - "Adsorption"	Store in suital	DOT label is 'Pr	Incineration,
Phenol	1E	VAPOR	Watercourse	Wear appropriate cloth	Stop the leaking source i	1E2 - "Adsorption"	Store in suital	DOT label is 'Pr	Incineration,
Phenol	2A	SOLID	On land	Wear appropriate cloth	Solid materials are usually	2A1 Normal weatl	Store in suital	DOT label is 'Pr	Incineration
Phenol	2A	SOLID	Watercourse	Wear appropriate cloth	"Booms", "Flow guides",	2A2 - "Skimmer", "	Store in suital	DOT label is 'Pr	Incineration,
Phenol	4N	SOLID	On land	Wear appropriate cloth	Stop the leaking source, .	4N1 "Pump into s	Store in suital	DOT label is 'Pr	Incineration
Phenol	4N	SOLID	Watercourse	Wear appropriate cloth	Certain chemicals in this g	4N2 Using "Skimr	Store in suital	DOT label is 'Pr	Incineration
Sodium hydroxide	4B	SOLID	On land	Wear appropriate cloth	Usually they are solid and	4B1 For liquid "Pr	Store in suital	DOT label is 'Ci	Discharge int
Sodium hydroxide	4B	SOLID	Watercourse	Wear appropriate cloth	For water soluble chemic	4B2 "Adsorption"	Store in suital	DOT label is 'Ci	Discharge int
Sulfuric acid	4A	LIQUID	On land	Wear appropriate cloth	Stop the leaking source	4A1 "Pump into s	Store in suital	DOT label is 'Ci	Add slowly to
Sulfuric acid	4A	LIQUID	Watercourse	Wear appropriate cloth	For water soluble chemic	4A2 "Adsorption"	Store in suital	DOT label is 'Ci	Add slowly to

Record 3

Table A1-2

Report from ERexpert

Countermesurement

30-Nov-95

Personal Protection	Containment/On land	Cleanup/On land	Containment/Watercours
Wear appropriate clothing, eg. nitrile gloves; chemical safety goggles; approved respirator; plastic apron, sleeves and boots, to prevent prolonged or repeated skin and eye contact. Employees should wash promptly when skin is wet or contaiminated. Remove clothing immediately if wet or contaminated to avoid flammability hazard.	Stop the leaking source, such as closing valve, pluging the leak. Using "Auxiliary tank", "Sump", "Dike", "Excavation" to confine the spilled material. If possible guide the spill to prepared surface to minimize the vertical seepage. If all the attempts failed, try reverse containment.	"Pump into suitable container" to recover spilled material. "Dredging" or "Shovelling" to remove the residuals (contaminated containment materials and bottom soil), or using in-place treatment by chemical or physical approaches as is appropriate.	Stop the leaking source if possible. Using "Booms", "Flow guides", "Curtains" to confine the spill in its most concentrated form.

Cleanup/Watercours	Storage	Shipping	Disposal Plan
"Skimmer", "Adsorption", "Gelling"	Store in tightly closed containers in a cool, well-ventilated area away from heat and ignition source. Protect from damage. Strore away from incompatible materials, such as strong oxidizers (eg. Chlorine, Bromine, and Flourine). Used ground drums and bond transfer containers equipped with self-closing valves, pressure vacuum bungs, and flame arresters. Use only non-sparking tools and equipment especially when opening and closing containers.	DOT label is "Flammable Liquid". The limit on passenger aircraft or railcar is 5 liters; on cargo aircraft shipment is 60 liters. It falls in DOT Hazard Class 3 and Packing Group II.	Follow all federal, provincial regulations for disposal. Use only licensed disposal and waste hauling companies. Disposal of small amounts of spilled material may be handled as described under "Leak or Spill Cleanup". Large spill must be delt with separately and must be handled by qualified disposal companies.

Figure A1-6

Toxicity : First Aid

Comp Name: Benzene

LC50: 13,228/4H (ihl-rat)

LD50: 3306 mg/kg (orl-rat)

Exposure Limits: TLV-TWA: 30 mg/m; TLV-STEL: 75 mg/m

Route of Entry: skin contact(irritates); skin absorbtion(readily absorbed, toxic); eye contact(irritates); inhalation(toxic, irritates); ingestion(toxic)

Effects of Acute Exposure: This product acts on the nerve system. It causes drowsiness, dizziness, headache, light-headedness, nausea and decreased coordination. It causes mild irritation of respiratory tract. High exposure may cause decreased judement, general malaise, lose balance, ringing in the ears or even unconsciousness and death. Skin contact causes degreasing which may cause irritation, inflammation, blistering, drying and scaling. Ingestion of small amount may cause dizziness, excitation, weakness, headache, breathlessness, constriction in the chest and pallor followed by flushing

Effects of Chronic Exposure: Benzene may impair the formation of red and white blood cells and platelets. Prolonged low level exposure can damage the nerve system characterized by impairment of hearing, chronic headache, dizziness, fatigue, visual disturbances, decreased balance and pallor. Repeated or prolonged skin contact can cause inflammation, dry scaling and blisters. Prolonged exposure may cause lesions resembling first and second degree burns.

Irritancy: Standard Draize Test: skn, rbt 30 mg/24H-moderate; eye, rbt 2mg/24H-severe

Potential Exposure: Benzene is used as constituent in motor fuels, as a solvent for fats, inks, oils, paints, plastics, and rubber, in the extraction of oils from seeds and nuts, and in photogravure printing. It is also used as a chemical intermediate. By alkylation, chlorination, nitration, and sulfonation, chemicals such as styrene, phenols, and maleic anhydride are produced. Benzene is also used in the manufacture of detergents, explosives, pharm. ceuticals, and dye-stuffs.

Points of attack: Blood; Central nervous system; Skin; Bone marrow; Eyes; Respiratory system.

Symptom: ACUTE: Strong CNS depressant, dizziness, weakness, euphoria, headache, irritability, nausea, vomiting, ataxia, staggering, anorexia, flatulence, tightness in chest, rapid irregular pulse, shallow and rapid respiration, ventricular irregularities, excitement, restless, delirium, blurred visual blindness, tremors, paralysis, coma, convulsions, massive hemorrhage of lungs, congestive gastritis, marked cerebral edema and kidney involvement. Death by cardiac and/or respiratory failure. LATE: Severe skin rash, capillary permeability, petechial hemorrhages, decreased

Treatment: EYE: Flush with lukewarm water for at least 15 min., holding the eyelids open. Take care not to rinse contaminated water into the non-affected eye. Seek medical attention. SKIN: Flush with lukewarm water for at least 15 min. Remove contaminated clothing, taking care not to spread the chemical. If extensive, remove clothing under running water. Discard or decontaminate clothing under running water. Seek medical attention unless contact has been slight. INHALATION: Take proper precautions to ensure your own safety before attempting rescue. Remove source or move victim to fresh air. If breathing has been stopped, trained personnel should begin artificial respiration, or if the heart has stopped, CPR immediately. Seek medical attention. INGESTION: Never give anything by mouth if victim is rapidly

Remark: Poisoning occurs most commonly via inhalation of the vapor, although benzene can penetrate the skin and cause poisoning. There is great individual variation in the signs

Figure A1-7

Appendix 2: The Knowledge Domain Matrices (KDMs) used in the SPILLexpert project.

The following KDM shows the knowledge layer that represents knowledge subdomain for the selection of treatment site. Together with the KDMs illustrated in Figures 4.6 and 4.11, the three KDMs form the knowledge precursor for the **ERexpert** program. The actual knowledge base in the form of production rules (refer to Appendix 3) can be generated accordingly.

Conditions / Conclusions	1. Try to treat the spill in-situ even it is not containable.	2. Haul to remote site for futher treatment.	3. Treat in-situ with makeshift process.	4. Treat on-site, but off stream.	5. Haul to remote disposal site for temp. dispose.	6. Disposal on-site temorarily.	7. Dilute and disperse if no suitable treatment or disposal site	8. Dilute and disperse because of equipment and supply difficulties.	9. Dilute and disperse because of technic difficulties.
1. Spill containable	F								
2. Remote treatment site available		T	F	F	F		F		
3. In-situ treatment site available			T	T	f		F		
4. Hauling vehicle/equipment available		T	f	f		F		F	
5. Plan can executed within time limit	T	T	T	T		F			F
6. Applicability of the plan				F					F
7. Field treatment method(s)/supplies available	T		T	T		f		F	
8. Disposal site available					T	T	F	F	F

Appendix 3: The current version of KBF used by the SPILLexpert project

FIND
CategoryID, ContTech, TrmtSite,

RULE 1
IF Variable1=YES AND Variable2=YES AND Variable3=YES
THEN CategoryID=1A,
EXP "Chemicals in this category are Low lying/ Flammable/ Vapour chemicals ",

RULE 2
IF Variable1=YES AND Variable2=YES AND Variable3=NO
THEN CategoryID=1B,
EXP "Chemicals in this category are Low lying/ Nonflammable/Vapour chemicals ",

RULE 3
IF Variable1=YES AND Variable2=NO AND Variable3=NO
THEN CategoryID=1C,
EXP "Chemicals in this category are Lighter/ Nonflammable/ Vapour chemicals ",

RULE 4
IF Variable1=YES AND Variable2=NO AND Variable3=YES
THEN CategoryID=1D,
EXP "Chemicals in this category are Lighter/ Flammable/ Vapour chemicals ",

RULE 5
IF Variable1=NO AND Variable5=NO AND Variable7=YES
THEN CategoryID=2A,
EXP "Chemicals in this category are Floating/ Solid chemicals ",

RULE 6
IF Variable7=YES AND Variable5=YES AND Variable3=YES AND Variable6=YES
THEN CategoryID=2B,
EXP "Chemicals in this category are Floating/ Flammable/ Volatile/Liquid chemicals.";

RULE 7
IF Variable7=YES AND Variable5=YES AND Variable3=NO AND Variable6=NO
THEN CategoryID=2C;
EXP "Chemicals in this category are Floating/ Nonflammable/ Nonvolatile/Liquid chemicals.",

RULE 8
IF Variable5=YES AND Variable6=YES AND Variable7=YES AND Variable3=NO
THEN CategoryID=2D;
EXP "Chemicals in this category are Floating/ Volatile/ Nonflammable/Liquid chemicals ",

RULE 9
IF Variable3=YES AND Variable5=YES AND Variable6=NO AND Variable7=YES
THEN CategoryID=2E;
EXP "Chemicals in this category are Floating/ Flammable/ Nonvolatile/Liquid chemicals ",

RULE 10
IF Variable1=NO AND Variable5=NO AND Variable7=NO
THEN CategoryID=3A;
EXP "Chemicals in this category are Sinking/Solid chemicals ";

RULE 11
IF Variable5=YES AND Variable7=NO
THEN CategoryID=3B,
EXP "Chemicals in this category are Sinking, Liquid chemicals ",

RULE 12
IF Variable4=YES AND Variable8=YES AND Variable11=NO
THEN CategoryID=4A,
EXP "Chemicals in this category are water soluble acids ",

RULE 13
IF Variable4=YES AND Variable8=YES AND Variable11=YES
THEN CategoryID=4A,
EXP "Chemicals in the category can dissolve and react with water to produce acids, can be treated as soluble acids"

RULE 14
IF Variable4=YES AND Variable8=NO AND Variable11=NO
THEN CategoryID=4B,
EXP "Chemicals in this category are water soluble bases ",

RULE 15
IF Variable4=YES AND Variable8=NO AND Variable11=YES
THEN CategoryID=4B,
EXP "Chemicals in this category can dissolve and react with water to produce bases They can be treated similarly as soluble bases compounds ",

RULE 16
IF Variable4=YES AND Variable9=YES
THEN CategoryID=4C,
EXP "Chemicals in this category are water soluble salts containing heavy metal ions ",

RULE 17
IF Variable4=YES AND Variable9=NO
THEN CategoryID=4D,
EXP "Chemicals in this category are water soluble salts that do not contain heavy metal cations ",

RULE 18
IF Variable4=YES AND Variable10=YES
THEN CategoryID=4E,
EXP "Chemicals in the category are water soluble and biodegradable chemicals ",

RULE 19
IF Variable4=YES AND Variable10=NO
THEN CategoryID=4F,
EXP "Chemicals in this category are water soluble and nonbiodegradable chemicals ".

RULE 21
IF Variable21=YES
THEN ContTech=Correct mechanic failure Run again,
EXP "Correction of the mechanic failure first by closing valve, clog leaks, or shut down pump, etc Then try to cleanup the spill You should first decide what type of chemical has been spilled (find Category ID), then run this part again "

RULE 28
IF Variable31=YES AND Variable32=YES AND Variable29= YES OR Variable30=YES
THEN ContTech=Self contained/Shovel/ Vacuuming;
EXP "Solid chemicals are usually self contained Recovery by shoveling or vacuuming ",

RULE 44
IF Variable43=YES AND Variable50=YES AND Variable37= YES OR Variable30=YES
THEN ContTech=Dike/Pump/Temp burial,
EXP "Natural dike or excavation to contain spilled sinking chemicals Pumping to recover the lost into auxiliary tanks or burial to temp contain the spill for later treatment ",
RULE 30
IF Variable31=YES AND Variable32=YES AND Variable42= NO AND Variable44=YES OR Variable45=YES
THEN ContTech=Treat as Category 2A/3A on dry surface,
EXP "Treat as Category 2A/3A chemicals on dry surface ",

RULE 22
IF Variable22=YES AND Variable27=YES
THEN ContTech=No spark/Inert foam cover/ Mist knock down,
EXP "Avoid spark Inert foam coverage, mist knock down ",

RULE 23
IF Variable26=YES AND Variable27=YES AND Variable22= YES OR Variable25=YES
THEN ContTech=No spark/Cryogenic con- densation/Fans/ Blowers,
EXP "Avoid spark Cryogenic condensation, or air dilution using fans or blowers ",

RULE 24
IF Variable22=YES AND Variable27=YES AND Variable28= YES
THEN ContTech=Avoid spark and encapsula- tion,

RULE 25
IF Variable23=YES AND Variable27=YES
THEN ContTech=Inert foam cover/Mist knock down,
EXP "Inert foam coverage with mist knock down ",

RULE 26
IF Variable26=YES AND Variable27=YES AND Variable23= YES OR Variable24=YES
THEN ContTech=Cryogenic condensation/ Fans/Blowers,
EXP "Cryogenic condensation, air dilution using fans or blowers ",

RULE 27
IF Variable23=YES AND Variable27=YES AND Variable28= YES
THEN ContTech=Encapsulation,

RULE 29
IF Variable31=YES AND Variable32=NO AND Variable29= YES OR Variable30=YES
THEN ContTech=Treat as liquid, prevnting run-off and assign liquid ID,
EXP "Fluidized solids shall treat as liquids Major concern is to prevent it run off into watercourse or sewage system. Response methods may be obtained by assign a proper liquid CategoryID to the target ",

RULE 79
IF Variable31=YES AND Variable32=YES AND Variable42= NO AND Variable46=YES OR Variable47=YES
THEN ContTech=Treat as Category 2A/3A on dry surface.

RULE 80
IF Variable31=YES AND Variable32=YES AND Variable42= NO AND Variable48=YES OR Variable49=YES
THEN ContTech=Treat as Category 2A/3A on dry surface,

RULE 31
IF Variable31=YES AND Variable32=NO AND Variable42= NO AND Variable44=YES OR Variable45=YES
THEN ContTech=Treat as fluidized Category 2A/3A chemicals;
EXP "Treat as fluidized Category 2A/3A chemicals "

RULE 81
IF Variable31=YES AND Variable32=NO AND Variable42= NO AND Variable46=YES OR Variable47=YES
THEN ContTech=Treat as fluidized Category 2A/3A chemicals,

RULE 82
IF Variable31=YES AND Variable32=NO AND Variable42= NO AND Variable48=YES OR Variable49=YES
THEN ContTech=Treat as fluidized Category 2A/3A chemicals,

RULE 32
IF Variable31=YES AND Variable33=YES AND Variable39= YES
THEN ContTech=No spark/Cover/Dike/Pump/Remove/ Burial;
EXP "Avoid spark Inert foam or plastic sheet coverage Earthen dike, trench, or excavation to contain the spill. Using pump or vacuum to collect the spill into auxiliary tank or sump, then using earth moving equipment to remove the bottom land. If possible direct spill to paved surface or use soil surface sealant to slow the vertical penetration. If no equipment available, try burial to temporarily contain the spill for later treatment.",

RULE 33
IF Variable31=YES AND Variable34=YES AND Variable39= YES
THEN ContTech=Dike/Pump/Remove bottom/ Temp burial,
EXP "Earthen dike, trench, or excavation to contain the spill. Using pump or vacuum to collect the spill into auxiliary tank or sump, then using earth moving equipment to remove the bottom land. If possible, direct spill to paved surface or use soil surface sealant to slow the vertical penetration. If no equipment available, try burial to temporarily contain the spill for later treatment ";

RULE 34
IF Variable31=YES AND Variable35=YES AND Variable39= YES
THEN ContTech=Cover/Dike/Pump/Remove bottom/Temp burial;
EXP "Inert foam or plastic sheet coverage Earthen dike, trench, or excavation to contain the spill. Using pump or vacuum to collect the spill into auxiliary tank or sump, then using earth moving equipment to remove the bottom land If possible, direct spill to paved surface or use soil surface sealant to slow the vertical penetration If no equipment available, try burial to temporarily contain the spill for later treatment.";

RULE 35
IF Variable31=YES AND Variable36=YES AND Variable39= YES

THEN ContTech=No spark/Dike/Pump/ Remove bottom/ Temp burial;
EXP "Avoid spark contact Earthen dike, trench, or excavation to contain the spill Using pump or vacuum to collect the spill into auxiliary tank or sump, then using earth moving equipment to remove the bottom land If possible, direct spill to paved surface or use soil surface sealant to slow the vertical penetration If no equipment available, try burial to temporarily contain the spill for later treatment ",

RULE 36
IF Variable31=YES AND Variable43=YES AND Variable48= YES AND Variable49=NO
THEN ContTech=No spark/Cover/Dike/Pump/ Temp burial,
EXP "Avoid spark Inert foam or plastic sheet coverage Build foamed polyurethane or concrete dike to contain the spill. Using pump or vacuum to collect the spill into auxiliary tank or sump If needed, improve the bottom condition by direct spill to paved surface or modify using surface sealant to slow vertical penetration If no equipment available, try burial to temporarily contain the spill for later treatment ",

RULE 37
IF Variable31=YES AND Variable34=YES AND VAriable38= YES AND Variable39=NO
THEN ContTech=Dike/Pump/Temp burial,
EXP "Build foamed polyurethane or concrete dike to contain the spill. Pump or vacuum to collect the spill into auxiliary tank or sump If needed, improve the bottom condition by direct spill to paved surface or modify using surface sealant to slow vertical penetration If no equipment available, try burial to temp contain the spill for later treatment ",

RULE 38
IF Variable31=YES AND Variable35=YES AND Variable38= YES AND Variable39=NO
THEN ContTech=Cover/Dike/Pump/Temp. burial;
EXP "Inert foam or plastic sheet coverage Build foamed polyurethane or concrete dike to contain the spill Using pump or vacuum to collect the spill into auxiliary tank or sump If needed, improve the bottom condition by direct spill to paved surface or modify using surface sealant to slow vertical penetration If no equipment available, try burial to temp contain the spill for later treatment ",

RULE 39
IF Variable31=YES AND VAriable36=YES AND Variable38= YES AND Variable39=NO
THEN ContTech=No spark/Dike/Pump/Temp burial,
EXP "Avoid spark contact Build foamed polyurethane or concrete dike to contain the spill Using pump or vacuum to collect the spill into auxiliary tank or sump If needed, improve the bottom condition by direct spill to paved surface or modify using surface sealant to slow vertical penetration If no equipment available, try burial to temporarily contain the spill for later treatment.",

RULE 40
IF Variable31=YES AND Variable37=YES AND Variable40= YES AND Variable41=YES
THEN ContTech=Treat as Category 2B chemicals,

RULE 41
IF Variable31=YES AND Variable37=YES AND Variable40= NO AND Variable41=NO
THEN ContTech=Treat as Category 2C chemicals,

RULE 42
IF Variable31=YES AND Variable37=YES AND Variable40= NO AND Var.able41=YES
THEN ContTech=22Treat as Category 2D chemicals,

RULE 43
IF Variable31=YES AND Variable37=YES AND Variable40= YES AND Variable41=NO
THEN ContTech=Treat as Category 2E chemicals,

RULE 45
IF Variable43=YES AND Variable39=YES AND Variable37= YES OR Variable30=YES
THEN ContTech=Under water dikes/Remove bottom by dredging/ Temp burial,
EXP "Construction of under water dikes, excavations and remove by dredging or burial to temporarily contain the spill for later treatment ",

RULE 46
IF Variable43=YES AND Variable27=YES AND Variable37= YES OR Variable30=YES
THEN ContTech=Curtain barriers/Pump,
EXP "Curtain barriers to contain the spill and pumping to recover the lost ",

RULE 47
IF Variable43=YES AND Variable27=YES AND Variable33 YES OR Variable36=YES
THEN ContTech=No spark/Booms/Skimmers, vacuuming,
EXP "Avoid spark Using booms, weirs, pneumatic barriers to contain the spill Recovery by using skimmers or vacuum collector into auxiliary tank ",

RULE 48
IF Variable43=YES AND Variable27=YES AND Variable34= YES OR Variable29=YES
THEN ContTech=Booms/Skimmers, vacuum,
EXP "Using booms/weirs/pneumatic barriers to contain the spill Recovery using skimmers/ vacuum collector into auxiliary tank ",

RULE 83
IF Variable27=YES AND Variable35=YES AND Variable43 YES
THEN ContTech=Booms/Skimmers, vacuum,

RULE 49
IF Variable43=YES AND Variable51=YES AND Variable33 YES OR Variable36=YES
THEN ContTech=No effective method/ Burning
EXP "No effective method to contain the spill Try burning the spill to cleanup the accident ",

RULE 50
IF Variable43=YES AND Variable34=YES OR Variable35 YES
THEN ContTech=No effective method/Water dilution/Biodegrad ,
EXP "No effective method to contain/cleanup floating chemicals Water dilution/leave for biodegradation process "

RULE 84
IF Variable43=YES AND Variable29=YES
THEN ContTech=No effective method/Water dilution/ Biodegrad ,

RULE 51
IF Variable43=YES AND Variable52=YES AND Variable33= YES OR Variable34=YES

THEN ContTech=Pump/Centrifuge separation/ Solvent extraction,
EXP "For removable volume of insoluble chemicals, first pump to auxiliary tank or sump Then using centrifuge separation or solvent extraction to treat the spill ",

RULE 85
IF Variable43=YES AND Variable52=YES AND Variable29=YES OR Variable30=YES
THEN ContTech=Pump/Centrifuge separation/Solvent extraction,

RULE 86
IF Variable43=YES AND Variable52=YES AND Variable35=YES OR Variable36=YES
THEN ContTech=Pump/Centrifuge separation/Solvent extraction,

RULE 87
IF Variable37=YES AND Variable43=YES AND Variable52=YES
THEN ContTech=Pump/Centrifuge separation/Solvent extraction,

RULE 52
IF Variable31=YES AND Variable40=YES AND Variable41=YES AND Variable42=YES AND Variable44=YES OR Variable45=YES
THEN ContTech=Treat as Category 2B chemicals,

RULE 88
IF Variable31=YES AND Variable40=YES AND Variable41=YES AND Variable42=YES AND Variable46=YES OR Variable47=YES
THEN ContTech=Treat as Category 2B chemicals,

RULE 89
IF Variable31=YES AND Variable40=YES AND Variable41=YES AND Variable42=YES AND Variable48=YES OR Variable49=YES
THEN ContTech=Treat as Category 2B chemicals,

RULE 53
IF Variable31=YES AND Variable40=NO AND Variable41=NO AND Variable42=YES AND Variable44=YES OR Variable45=YES
THEN ContTech=Treat as Category 2C chemicals;

RULE 90
IF Variable31=YES AND Variable40=NO AND Variable41=NO AND Variable42=YES AND Variable46=YES OR Variable47=YES
THEN ContTech=Treat as Category 2C chemicals,

RULE 91
IF Variable31=YES AND Variable40=NO AND Variable41=NO AND Variable42=YES AND Variable48=YES OR Variable49=YES
THEN ContTech=Treat as Category 2C chemicals,

RULE 54
IF Variable31=YES AND Variable40=NO AND Variable41=YES AND Variable42=YES AND Variable44=YES OR Variable45=YES
THEN ContTech=Treat as Category 2D chemicals,

RULE 92
IF Variable31=YES AND Variable40=NO AND Variable41=YES AND Variable42=YES AND Variable46=YES OR Variable47=YES
THEN ContTech=Treat as Category 2D chemicals,

RULE 93
IF Variable31=YES AND Variable40=NO AND Variable41=YES AND Variable42=YES AND Variable48=YES OR Variable49=YES
THEN ContTech=Treat as Category 2D chemicals,

RULE 55
IF Variable31=YES AND Variable40=YES AND Variable41=NO AND Variable42=YES AND Variable44=YES OR Variable45=YES
THEN ContTech=Treat as Category 2E chemicals,

RULE 94
IF Variable31=YES AND Variable40=YES AND Variable41=NO AND Variable42=YES AND Variable46=YES OR Variable47=YES
THEN ContTech=Treat as Category 2E chemicals,

RULE 95
IF Variable31=YES AND Variable40=YES AND Variable41=NO AND Variable42=YES AND Variable48=YES OR Variable49=YES
THEN ContTech=Treat as Category 2E chemicals,

RULE 56
IF Variable43=YES AND Variable44=YES OR Variable45=YES
THEN ContTech=No effective method/Water dilution/Biodegrad.,
EXP "No effective containment method for dissolving chemicals in large, open watercourse The only effective method is to disperse by water dilution into natural environment. Some chemicals may undergo biodegradation.",

RULE 96
IF Variable43=YES AND Variable46=YES OR Variable47=YES
THEN ContTech=No effective method/Water dilution/ Biodegrad ,

RULE 97
IF Variable43=YES AND Variable48=YES OR Variable49=YES
THEN ContTech=No effective method/Water dilution/ Biodegrad.;

RULE 57
IF Variable43=YES AND Variable52=YES AND Variable44=YES OR Variable45=YES
THEN ContTech=Pump/Solvent extraction/ Gelation;
EXP "For removable volume of soluble chemicals (4A-4F), first pump to auxiliary tank or sump, then using solvent extraction or gelation to separate the spill from water body ",

RULE 98
IF Variable43=YES AND Variable52=YES AND Variable46=YES OR Variable47=YES
THEN ContTech=Pump/Solvent extraction/ Gelatation,

RULE 99
IF Variable43=YES AND Variable52=YES AND Variable48=YES OR Variable49=YES
THEN ContTech=Pump/Solvant extraction/ Gelatation,

RULE 58
IF Variable27=YES AND Variable28=YES AND Variabl43= YES AND Variabl44=YES
THEN ContTech=Sealed boom/Neutralization/Adsorption/ Ion exhange,
EXP "Using sealed booms to contain spill in limited, calm water area. Neutralization with Na_2CO_3, $NaHCO_3$, $Ca(OH)_2$, CaO, $CaCO_3$, etc, except hydrogen cyanide Adsorption using activated carbon, or ion exchange using process resin.";

RULE 59
IF Variable28=YES AND Variable43=YES AND Variable53= YES AND Variable44=YES
THEN ContTech=Diversion/Neutralization/ Adsorption/Ion exchange;
EXP "For limited spill size and flowing water, try diversion of contaminated water or reverse into separated area. Then using neutralization with Na_2CO_3, $NaHCO_3$, $Ca(OH)_2$, CaO, $CaCO_3$, etc, except hydrogen cyanide, adsorption using activated carbon, or ion exchange using process resin ",

RULE 60
IF Variable27=YES AND Variable28=YES AND Variable43= YES AND Variable45=YES
THEN ContTech=Sealed boom/Neutralization/Adsorption/ Ion exchange;
EXP "Using sealed booms to contain spill in limited, calm water. Neutralization with $NaHCO_3$ or CO_2. Adsorption using activated carbon, or ion exchange using process resin ";

RULE 61
IF Variable28=YES AND Variable43=YES AND Variable53= YES AND Variable45=YES
THEN ContTech=Diversion/Neutralization/ Adsorption/Ion exchange,
EXP "For limited size and flowing water, try diversion of contaminated water, reverse into separated area Then using neutralization with $NaHCO_3$, CO_2, etc., adsorption using activated carbon, or ion exchange using process resin.";

RULE 62
IF Variable27=YES AND Variable28=YES AND Variable43= YES AND Variable46=YES
THEN ContTech=Sealed boom/Precipitation/ Adsorption /Ion exchange;
EXP "Using sealed booms to contain spill in limited, calm water area. Precipitation with $CaSO_4$, $CaCl_2$, $CaCO_3$, or Na_2S. Adsorption using activated carbon, or ion exchange using process resin ",

RULE 63
IF Variable28=YES AND Variable43=YES AND Variable53= YES AND Variable46=YES
THEN ContTech=Diversion/Precipitation/ Adsorption/Ion exchange;
EXP "For limited spill size and flowing water, try diversion of contaminated water or reverse into separated area Then using precipitation with $CaSO_4$, $CaCO_3$, $CaCl_2$, or Na_2S, etc., adsorption using activated carbon, or ion exchange using process resin ";

RULE 64
IF Variable27=YES AND Variable28=YES AND Variable43= YES AND Variable47=YES
THEN ContTech=Sealed boom/Oxidation/ Adsorption /Ion exchange;
EXP "Using sealed booms to contain spill Oxidation using H_2O_2 or O_3 for cyanide salt, adsorption using activated carbon, or ion exchange using process resin for other salts in this group ",

RULE 65
IF Variable28=YES AND Variable43=YES AND Variable53= YES AND Variable47=YES
THEN ContTech=Diversion/Oxidation/ Adsorp- tion/Ion exchange,
EXP "For limited spill size and flowing water, try diversion of contaminated water or reverse into separated area Oxidation using H_2O_2 or O_3 for cyanide salt, adsorption using activated carbon, or ion exchange using process resin for other salts in this group ",

RULE 66
IF Variable27=YES ADN Variable28=YES AND Variable43= YES AND Variable48=YES
THEN ContTech=Sealed boom/Adsorption/ Ion exchange,
EXP "Using sealed booms to contain spill Adsorption using activated carbon or polyurethane foam, ion exchange using process resin ";

RULE 67
IF Variable28=YES AND Variable43=YES AND Variable53= YES AND Variable48=YES
THEN ContTech=Diversion/Adsorption/Ion exchange,
EXP "For limited spill size and flowing water, try diversion of contaminated water or reverse into a separated area Adsorption using activated carbon or polyurethane foam, ion exchange using process resin ".

RULE 68
IF Variable27=YES AND Variable28=YES AND Variable43= YES AND Variable49=YES
THEN ContTech=Sealed boom/Adsorption/ Ion exchange,
EXP "Using sealed booms to contain spill Adsorption using activated carbon or polyurethane foam, ion exchange using process resin ",

RULE 69
IF Variable28=YES AND Variable43=YES AND Variable53= YES AND Variable49=YES
THEN ContTech=Diversion/Adsorption/Ion exchange,
EXP "For limited spill size and flowing water, try diversion of contaminated water or reverse into separated area Adsorption using activated carbon or polyurethane foam, ion exchange using process resin ",

RULE 70
IF Variable60=NO AND Variable64=YES AND Variable66= YES
THEN TrmtSite=Treat on site,
EXP "Even the spill is not containable, once the treatment methods and supplies are available, try to treat in field ".

RULE 71
IF Variable61=YES AND Variable63=YES AND Variable66= YES
THEN TrmtSite=Remote treatment,
EXP "Haul to remote site for further treatment This is always the preference if possible ".

RULE 72
IF Variable62=YES AND Variable64=YES AND Variable66= YES AND Variable65=YES
AND Variable61=NO OR Variable63=NO
THEN TrmtSite=In-situ treatment,
EXP "In-situ treatment with a makeshift process ".

RULE 73
IF Variable62=YES AND Variable64=YES AND Variable66=YES AND Variable65=NO
AND Variable61=NO OR Variable63=NO
THEN TrmtSite=On site, off stream treatment;

RULE 74
IF Variable67=YES AND Variable61=NO OR Variable62=NO
THEN TrmtSite=Disposal,
EXP "Since no treatment sites are available at the moment, haul to a disposal site for disposal, when suitable further treatment can be resumed ",

RULE 75
IF Variable67=YES AND Variable66=NO AND Variable63=NO OR Variable64=NO
THEN TrmtSite=Disposal on-site temporarily,
EXP "Since no hauling facility or field treatment supplies can be arranged within time limit Temporarily disposal nearby and if later hauling facility become available, haul to disposal site for disposal Whenever suitable further treatment can be resumed ",

RULE 76
IF Variable61=NO AND Variable62=NO AND Variable67=NO
THEN TrmtSite=Dilute and disperse,
EXP "If no treatment and disposal sites available, and the chemical can be diluted with water, dilute and disperse to natural environment seems to be the last choice.",

RULE 77
IF Variable63=NO AND Variable64=NO AND Variable67=NO
THEN TrmtSite=Dilute and disperse,
EXP "If treatment supplies or hauling facilities are not available, and the chemical(s) can be diluted with water, dilute and disperse to natural environment can to be an alternative However, if other means of treatment possible, i e. Temporary burial, dilute and disperse is always the last option, especially when toxic chemicals are involved ",

RULE 78
IF Variable65=NO AND Variable66=NO AND Variable67=NO
THEN TrmtSite=Dilute and disperse;
EXP "If other means of treatment possible, i.e. temporary on-site burial, dilute and disperse is always the last option, especially when facing with highly toxic chemicals. You can always hope that technology will progress that the impossible today maybe possible tomorrow ",

ASK Variable1 "Vapor chemicals?" OPTION YES, NO;
ASK Variable2 "Vapor density greater than air?" OPTION YES, NO,
ASK Variable3 "Flammable chemicals?" OPTION YES, NO,
ASK Variable4 "Water soluble chemicals?" OPTION YES, NO,
ASK Variable5 "Liquid chemicals?" OPTION YES, NO;
ASK Variable6 "Volatile chemicals?" OPTION YES, NO;
ASK Variable7 "Density (liquid/solid) less than water?" OPTION YES, NO,
ASK Variable8 "Acidic chemicals?" OPTION YES, NO;
ASK Variable9 "Salts contain heavy metal ions?" OPTION YES, NO,
ASK Variable10 "Biodegradable chemicals?" OPTION YES, NO,
ASK Variable11 "React with water?" OPTION YES, NO;
ASK Variable21 "Mechanic failure?" OPTION YES, NO,
ASK Variable31 "Spill on land?" OPTION YES, NO,
ASK Variable42 "Liquid state?" OPTION YES, NO,
ASK Variable22 "Category 1A chemicals?" OPTION YES, NO;
ASK Variable23 "Category 1B chemicals?" OPTION YES, NO;
ASK Variable24 "Category 1C chemicals?" OPTION YES, NO;
ASK Variable25 "Category 1D chemicals?" OPTION YES, NO;
ASK Variable29 "Category 2A chemicals?" OPTION YES, NO,
ASK Variable26 "Sheltered area?" OPTION YES, NO;
ASK Variable27 "Calm area?" OPTION YES, NO,
ASK Variable28 "Limited spill size?" OPTION YES, NO;
ASK Variable43 "Spill in watercourse?" OPTION YES, NO,
ASK Variable30 "Category 3A chemicals?" OPTION YES, NO;
ASK Variable32 "Dry ground surface?" OPTION YES, NO;
ASK Variable33 "Category 2B chemicals?" OPTION YES, NO;
ASK Variable34 "Category 2C chemicals?" OPTION YES, NO;
ASK Variable35 "Category 2D chemicals?" OPTION YES, NO;
ASK Variable36 "Category 2E chemicals?" OPTION YES, NO;
ASK Variable37 "Category 3B chemicals?" OPTION YES, NO;
ASK Variable38 "Flat surface?" OPTION YES, NO;
ASK Variable39 "Soft surface/bottom?" OPTION YES, NO;
ASK Variable40 "Flammable?" OPTION YES, NO;
ASK Variable41 "Volatile?" OPTION YES, NO,
ASK Variable44 "Category 4A chemicals?" OPTION YES, NO,
ASK Variable45 "Category 4B chemicals?" OPTION YES, NO;
ASK Variable46 "Category 4C chemicals?" OPTION YES, NO,
ASK Variable47 "Category 4D chemicals?" OPTION YES, NO;
ASK Variable48 "Category 4E chemicals?" OPTION YES, NO,
ASK Variable49 "Category 4F chemicals?" OPTION YES, NO;
ASK Variable50 "Natural barrier exists?" OPTION YES, NO;
ASK Variable51 "Geographically right area for burning?" OPTION YES, NO;
ASK Variable52 "Small removable volume?" OPTION YES, NO,
ASK Variable53 "Clear area?" OPTION YES, NO;
ASK Variable60 "Is spill containable?" OPTION YES, NO,
ASK Variable61 "Is remote treatment site available?" OPTION YES, NO;
ASK Variable62 "Is field treatment site available?" OPTION YES, NO;
ASK Variable63 "Are hauling vehicle and equipment available?" OPTION YES, NO;
ASK Variable64 "Are field treatment methods and supplies available?" OPTION YES, NO,
ASK Variable65 "Applicability of the plan?" OPTION YES, NO;
ASK Variable66 "Can the plan been executed within time mit?" OPTION YES, NO;
ASK Variable67 "Is disposal site available?" OPTION YES, NO;

Appendix 4: Principle theory of quadrupole ion trap mass spectrometry

(1) The geometry of an ion-trap device is governed by the following equation: $r_0^2 = 2z_0^2$

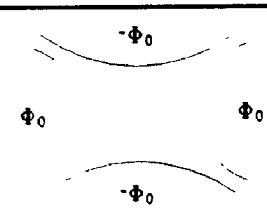
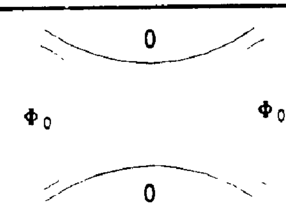
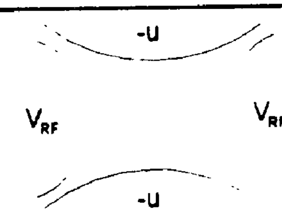
r_0 --- radius of ring electrode;

$2z_0$ --- separation between two end-cap electrodes.

(2) The Mathieu second-order differential equation: $\frac{d^2\mu}{d^2\xi} + (a_\mu - 2q_\mu \cos 2\xi)\mu = 0$

μ --- radial and axial coordinate axes (r, z);
ξ --- dimensionless quantity, equal to $\frac{\Omega t}{2}$;
Ω --- radial frequency of the alternating component of potential Φ_0.

(3) Different modes of operation of an ion-trap device is summarized in the following table:

	Mode I	Mode II	Mode III
Operational mode			
E /end-cap electrodes	$-\Phi_0 = -(U + V\cos\Omega t)$	0	-U
E /ring electrodes	$\Phi_0 = U + V\cos\Omega t$	$\Phi_0 = U + V\cos\Omega t$	$V_{RF} = V\cos\Omega t$
Field expression	$\Phi(r,t) = \frac{\Phi_0}{r_0^2}(r^2 - 2z^2)$	$\Phi(r,t) = \frac{\Phi_0}{2r_0^2}(r^2 - 2z^2) + \frac{\Phi_0}{2}$	$\Phi(r,t) = \frac{U + V\cos\Omega t}{2r_0^2}(r^2 - 2z^2) + \frac{V\cos\Omega t - U}{2}$
Stability parameters			
a_r	$8eU/mr_0^2\Omega^2$	$4eU/mr_0^2\Omega^2$	$4eU/mr_0^2\Omega^2$
q_r	$4eV/mr_0^2\Omega^2$	$2eV/mr_0^2\Omega^2$	$2eV/mr_0^2\Omega^2$
a_z	$-16eU/mr_0^2\Omega^2$	$-8eU/mr_0^2\Omega^2$	$-8eU/mr_0^2\Omega^2$
q_z	$-8eV/mr_0^2\Omega^2$	$-4eV/mr_0^2\Omega^2$	$-4eV/mr_0^2\Omega^2$

Appendix 5: The current version of KBF used by GCMSdiagnosis

FIND
Leaking, Injection, Column, IonTrap, CarrierGas, Sample, Communication,

RULE 1
IF Variable1=YES OR Variable2=YES AND Variable8=YES AND Variable22=NO
THEN Leaking=Contaminated GC eluting,
EXP "If IT < 5,000 after a ballistically temperature increase of the GC zone that indicates a contaminated GC eluting To locate a single source, run again and be more specific ",

RULE 2
IF Variable1=YES AND Variable5=YES AND Variable6=YES OR Variable28=YES
THEN Leaking=System leaking,
EXP "Leaking cause Air/Water gets into the system. Especially after a recent connection change Need to completely purge pneumatic lines ",

RULE 3
IF Variable10=YES AND Variable5=YES AND Variable6=YES AND Variable18=YES
THEN Leaking=System system leaking (severe),
EXP "Leaking cause Air/Water gets into the system. Especially after a recent connection change Need to completely purge pneumatic lines ",

RULE 4
IF Variable10=YES AND Variable18=YES AND Variable23=YES
THEN Leaking=Severe leaking at col./septum,
EXP "Check column connections, tighten the nuts Sometimes, change septum in the injector port may be effective ",

RULE 5
IF Variable1=YES AND Variable10=YES AND Variable28=YES
THEN Leaking=Leaking at col /septum,
EXP "Check column connections, tighten the nuts. Sometimes, change septum in the injector port may be effective ",

RULE 6
IF Variable10=YES AND Variable19=YES OR Variable20=YES AND Variable22=YES
THEN Injection=Inject port temp not hot enough;
EXP "Peak broadening observed, especially the high boiler components tailing server, the problem could cause by low column/injector temperature ",

RULE 7
IF Variable10=YES AND Variable21=YES AND Variable18=NO
THEN Injection=Split ratio too high,
EXP "Reduce split ratio, or switch to splitless mode to confirm the cause of the problem ",

RULE 8
IF Variable25-YES AND Variable20=YES OR Variable24=YES AND Variable11=YES
THEN Injection=Injection speed too high,
EXP "Certain types of chemicals need suitable injection speed Adjust to 1 0 ul/sec ",

RULE 9
IF Variable10=YES AND Variable18=NO AND Variable23=YES
THEN Injection=Syringe defect./partially plugged,
EXP "If no leaking, and splitless mode is used, check syringe to see whether it is partially plugged.",

RULE 10
IF Variable1=YES OR Variable2=YES AND Variable3=NO AND Variable4=YES
THEN Injection=Septum bleeding/coring,
EXP "If peak with m/z = 207 observed, along with discrete ion clusters Change septum " ;

RULE 11
IF Variable10=YES AND Variable11=YES
THEN Injection=Improving injection using SPI,
EXP "Strong acidic or basic sample could cause sensitivity and/or resolution deterioration ";

RULE 12
IF Variable25=YES AND Variable11=YES
THEN Injection=Improving injection using SPI,
EXP "Strong acidic or basic sample could cause sensitivity and/or resolution deterioration ",

RULE 13
IF Variable1=YES OR Variable2=YES AND Variable3=YES AND Variable4=YES
THEN Column=Column bleeding,
EXP "If peak with m/z = 207 observed, along with ion clusters in the whole mass range. It could cause by column bleeding. Change column or reduce column temperature " ,

RULE 14
IF Variable1=YES AND Variable18=YES AND Variable20=YES
THEN Column=Column damaged, replace;
EXP "When leaking is not detected, this may suggest that the column is damaged and replacement is required ",

RULE 15
IF Variable25=YES AND Variable20=YES OR Variable24=YES OR Variable27=YES AND Variable18=YES
THEN Column=Column degraded, replace,
EXP "Band broadening includes tailing and/or leading peaks When leaking is not detected, this may suggest that the column is damaged and replacement is required ",

RULE 16
IF Variable25=YES AND Variable20=YES AND Variable17=YES
THEN Column=Col. temp./program not optimized,
EXP "Also check the injector port temperature ";

RULE 17
IF Variable25=YES AND Variable27=YES AND Variable20=YES
THEN Column=Column overload,
EXP "Usually, the column load is below 300 - 400 ng Reduce the column load by inject less volume or dilute sample solution and try again ",

RULE 18
IF Variable25=YES AND Variable19=YES AND Variable22=NO

THEN IonTrap=Mass transfer line temp low;
EXP "If increase temperature in the GC part of the instrument does not cure the problem, one possibility left is the temperature of the mass transfer line is set too low ",

RULE 19
IF Variable1=YES OR Variable2=NO AND Variable7=YES AND Variable8=YES
THEN IonTrap=Iontrap contam /condit overnight,
EXP "When ionization time is shortened, condition the ion-trap overnight at the temperature given by the manual Try again to see if any improvement.",

RULE 20
IF Variable1=YES AND Variable8=YES AND Variable9=YES AND Variable7=NO OR Variable2=NO
THEN IonTrap=Iontrap contam severe/cleaning,
EXP "When ionization time is severely shortened and conditioning doesn't help reducing the problem It means that the ion-trap is severely containminated and physical cleaning of the device is necessary.",

RULE 21
IF Variable10=YES AND Variable19=YES OR Variable20= YES AND Variable22=NO
THEN IonTrap=Iontrap temp. low,
EXP "It could cause by a low ion-trap temperature. Check and adjust the temperature to 200-250 C.",

RULE 22
IF Variable10=YES AND Variable25=YES AND Variable19= YES OR Variable20=YES
THEN IonTrap=Increase manifold temp.,
EXP "For low conc. (<=10 ng) sample analysis, set manifold temperature to 250-300 C.";

RULE 23
IF Variable25=YES AND Variable15=NO
THEN IonTrap=AGC target value high,
EXP "Tuning may required to adjust AGC value to 5000-20,000 range.";

RULE 24
IF Variable10=YES AND Variable12=YES AND Variable13= YES
THEN IonTrap=Wrong reagent gas,
EXP "Once aliphatic sample involved, methane is not a proper reagent gas ",

RULE 25
IF Variable10=YES AND Variable14=YES
THEN IonTrap= Ion neutralization reaction,
EXP "Ion neutralization reaction that can be sever if a proton donor or acceptor type of sample involved, corrective measure by setting target val. at/below 20,000 with a filament emission current of 20 μA or more ";

RULE 26
IF Variable10=YES AND Variable12=YES AND Variable15= YES AND Variable16=YES
THEN IonTrap=Poor scan function,
EXP "Check the scan function line by line Specifically, the r.f and axial modulation amplitude and frequency, BBISO design, time intervals, etc In reality, the details for each line of command is established and fine tuned in a series of experiments ",

RULE 27
IF Variable10=YES AND Variable9=NO AND Variable12= YES
THEN IonTrap=Poor scan function,
EXP "Check the scan function line by line Specifically, the r f. and axial modulation amplitude and frequency, BBISO design, time intervals, etc In reality, the details for each line of command is established and fine tuned in a series of experiments.",

RULE 28
IF Variable10=YES AND Variable12=YES OR Variable32 =YES
THEN IonTrap=Poor scan function,
EXP "Check the scan function line by line Specifically, the r f and axial modulation amplitude and frequency, BBISO design, time intervals, etc In reality, the details for each line of command is established and fine tuned in a series of experiments ",

RULE 29
IF Variable10=YES AND Variable15=NO
THEN IonTrap=Wrong AGC target value,
EXP "AGC target value too small (<5,000) Tuning may required to adjust AGC value to 5000-20 000 range ",

RULE 30
IF Variable10=YES AND Variable16=NO
THEN IonTrap=Filament current small,
EXP "Filament current too small (<20 μA), tuning may required to adjust to the range ",
RULE 31
IF Variable25=YES AND Variable26=YES
THEN IonTrap=Electron multiplier GAIN drop off,
EXP "Electron multiplier GAIN (10^5) drop off, tuning may required to adjust to the range ",

RULE 32
IF Variable25=YES AND Variable20=NO
THEN IonTrap=Incorrect AM amplitude,

RULE 33
IF Variable1=YES AND Variable17=YES AND Variable18= YES
THEN CarrierGas=Carrier gas flow high,
EXP "Carrier gas flow high (>2 0ml/min), adjust to 0 5-2 0 ml/min ',

RULE 34
IF Variable25=YES AND Variable17=YES AND Variable18= YES
THEN CarrierGas=Carrier gas flow high,
EXP "Carrier gas flow too high (>2 0ml/min), adjust to 0 5-2 0 ml/min ",

RULE 35
IF Variable25=YES AND Variable18=YES AND Variable24= NO
THEN CarrierGas=Unstable carrier gas flow,

RULE 36
IF Variable1=YES AND Variable5=YES AND Variable6=NO
THEN CarrierGas=Poor carrier gas quality,
EXP "Once oxygen peak presence, confirmed poor carrier gas quality ",

RULE 37
IF Variable1=YES AND Variable2=YES AND Variable5=NO OR Variable6=NO
THEN CarrierGas=Dirty plumbing,

RULE 38
IF Variable10=YES AND Variable23=YES AND Variable24=YES
THEN Sample=Active surface in injector or col ,
EXP "Sample being absorbed by active surface in injector or column ",

RULE 39
IF Variable1=YES AND Variable8=YES AND Variable7=NO
THEN Sample=Matrix effects;
EXP "Complex matrix without pretreatment, using SFE, SPE to eliminate matrix effect ";

RULE 40
IF Variable29=YES AND Variable31=YES
THEN Sample=Below detection limit,
EXP "Could below ion trap detection limit When analyzing pesticides or HOCs, use split mode and combine ECD to confirm the presence of trace analytes in sample ",

RULE 41
IF Variable29=YES AND Variable30=YES
THEN Communication=Com failure, detach and locate;
EXP "Detach suspected instrumental component to locate problem source ",

RULE 42
IF Variable29=YES AND Variable30=NO
THEN Communication=Com failure, check following,
EXP "Check following list 1) power off or fuse blown; sample injected into wrong column 2) syringe defective 3) oven not heated or col temp. too low 4) no carrier gas flow, col broken 5) injector part temp too low ",

RULE 43
IF Variable33=YES
THEN Communication=Com failure, check following,
EXP "Possible failure from IEEE serial communication board, IEEE cable, SAP PC board Reference SATURN manual V 1, Operator's 8-4 for settings, etc ",

RULE 44
IF Variable34=YES
THEN Communication=Com failure, check following,
EXP "Possible failure from GC serial communication PC board, GC CPU board, RS-232 cable, COM1 port Reference SATURN manual, V 1, Operator's 8-5 for PC board switch settings, etc ",

ASK Variable1 "High background noise masking sensitivity?"
OPTION YES, NO,
ASK Variable2 "Noise level decrease as column temp decrease?"
OPTION YES, NO,
ASK Variable3 "Ion cluster throughout the whole mass range?"
OPTION YES, NO,
ASK Variable4 "Silicone peaks at m/z = 207, 73, 281, 327, with intensity decrease in the order?"
OPTION YES, NO,
ASK Variable5 "Sever water peak indicated by HH2O(18)/HH3O(19) << 10 1?"
OPTION YES, NO,
ASK Variable6 "Air peaks indicated by m/z = 32, 28, 18 broadened?"
OPTION YES, NO,
ASK Variable7 "Conditioning suspected instrumental components effectively reduce problem?"
OPTION YES, NO,
ASK Variable8 "Short ionization time (<15,000 us)?"
OPTION YES, NO,
ASK Variable9 "RF tuning difficult?"
OPTION YES, NO,
ASK Variable10 "Low sensitivity?"
OPTION YES, NO;
ASK Variable11 "Acidic or basic polar sample?"
OPTION YES, NO,
ASK Variable12 "Weak daughter ion intensity?"
OPTION YES, NO,
ASK Variable13 "Chemical ionization mode?"
OPTION YES, NO;
ASK Variable14 "Significant (M+1) peak present?"
OPTION YES, NO,
ASK Variable15 "Normal AGC target value (between 5,000 to 20,000) ?"
OPTION YES, NO;
ASK Variable16 "Regular electron multiplier emission current (between 20 to 50 μA)?"
OPTION YES, NO,
ASK Variable17 "Unresolved peaks?"
OPTION YES, NO,
ASK Variable18 "Retention time change from run to run?"
OPTION YES, NO,
ASK Variable19 "High boiler sever tailing?"
OPTION YES, NO;
ASK Variable20 "Band broadening?"
OPTION YES, NO,
ASK Variable21 "Split injection mode?"
OPTION YES, NO,
ASK Variable22 "Increase GC zone temperature reduce the problem?"
OPTION YES, NO,
ASK Variable23 "Peak size changes from run to run?"
OPTION YES, NO,
ASK Variable24 "Peak tailing?"
OPTION YES, NO,
ASK Variable25 "Poor resolution?"
OPTION YES, NO,
ASK Variable26 "Varying ionization time?"
OPTION YES, NO;
ASK Variable27 "Leading peaks?"
OPTION YES, NO,
ASK Variable28 "Baseline drifting?"
OPTION YES, NO,
ASK Variable29 "No signal?"
OPTION YES, NO;
ASK Variable30 "Suspecting instrumental malfunction?"
OPTION YES, NO,
ASK Variable31 "Analyzing trace amount of sample?"
OPTION YES, NO;
ASK Variable32 "Parent ion still observed after tandem MS process?"
OPTION YES, NO,
ASK Variable33 "SATURN software could not start?"
OPTION YES, NO;
ASK Variable34 "Can't communicate with GC?"
OPTION YES, NO;